The Quantification of SIMS depth profiles by Maximum Entropy reconstruction.

Paul Allen BSc(Hons).

Thesis for the degree of

Doctor of Philosophy

submitted to

The University of Warwick

Department of Physics

February 1994.

Table of Contents.

List of Figures.

List of Tables.

Acknowledgements.

I would like to acknowledge the help of the following people during the course of this project.

Firstly, I would like to thank Dr R.Collins, for the original idea of using MaxEnt for SIMS depth profiles, and for his guidance during this research.

I would particularly like to thank my supervisor, Mark Dowsett, for his guidance and without whom this work would not have been possible.

All the members of the Advanced SIMS Projects group at Warwick for many useful discussions, and for making my time here so enjoyable, and particularly Bob Barlow for instructing me (sometimes repeatedly) on how to operate the EVA2000 instrument.

Thanks also to Prof. M. Maier, for providing the silicon-in-gallium arsenide samples used in chapter 6.

My parents for their support, both moral and financial, throughout my academic career. Finally my girlfriend Patti for, well, everything really.

I acknowledge the SERC and Kratos for the CASE studentship under which the majority of this work was performed.

Declaration.

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It contains an account of my work in the Department of Physics at the University of Warwick during the period October 1990 to February 1994, under the supervision of Dr. M G Dowsett.

No part of this thesis has been used previously in a degree thesis submitted to this or any other university. The work described is the result of my own research except where specifically acknowledged in the text.

P N Allen

February 1994.

List of Publications.

Deconvolution of Concentration profiles from SIMS data using measured response functions. R.Collins, M.G.Dowsett, and P.N.Allen. *Secondary Ion Mass Spectrometry, Proceedings of SIMS VIII*, p.111, Wiley, Chichester, (1992).

SIMS Profile Quantification by Maximum Entropy Deconvolution. P.N.Allen, M.G.Dowsett, and R.Collins. *Surf. Int. Anal.* **20**, 696, (1993).

Secondary ion mass spectrometry analysis of ultra-thin impurity layers in semiconductors and their use in quantification, instrumental assessment, and fundamental measurements. M.G.Dowsett, R.D.Barlow, and P.N.Allen. *Proc. 2nd Int. Wkshp on the Measurement and Characterisation of Ultra-Shallow Doping Profiles in Semiconductors*, eds. R.Subramanyan, C.Osborn, P.Rai-Choudhury, p.127 (MCNC, 1993), and *J. Vac. Sci. Technol. B.* **12(1)**, 1, (1994).

An Analytic Form for the SIMS Response Function Measured from Ultra-thin Impurity Layers. M.G.Dowsett, G.Rowlands, P.N.Allen, and R.D.Barlow, *J. Appl. Phys.* (in press).

Maximum Entropy quantification of SIMS depth profiles - behaviour as a function of primary ion energy. P.N.Allen, and M.G.Dowsett, *Surf. Int. Anal.* (in press).

Work on this project has been presented at SIMS VIII (Amsterdam, 1991), the SIMS Users Forum (Loughborough, 1992), QSA7 (Surrey, 1993), he 2nd Int. Workshop on Measurement and Characterisation of Ultra-Shallow Doping Profiles in Semiconductors (USA,1993) and the SIMS Users Forum Warwick, 1993).

Abstract.

The quantification procedures applied to raw SIMS data were devised on the basis of a simple model for the sputtering and ionisation that occur during measurement. The model and the associated quantification procedures have long been known to be inaccurate. If SIMS is to remain a useful analysis tool in the future, the quantification procedures must be adjusted such that current features of interest are accurately measured.

This thesis describes the development of a more accurate (though empirical) model for the effects of the analysis, using the convolution integral. We propose a method for the quantification of SIMS depth profiles appropriate to this model, using Maximum Entropy (MaxEnt) reconstruction. SIMS depth profile data differ significantly from previous applications of the MaxEnt method: the very high signal to background ratio of the technique has lead users to plot the results on a logarithmic axis, giving much importance to extremely small signals.

The noise on SIMS depth profiles has been characterised. A number of optimisation algorithms have been developed and tested, and the performance of the MaxEnt method on SIMS data has been assessed. A novel form of the entropy, particularly suited to SIMS depth profiles, has been suggested. This form has given excellent results.

1.0. Introduction.

1.1. Introduction to the problem.

Secondary Ion Mass Spectrometry (SIMS) is a highly sensitive mass analysis tool. Although the technique excels in a number of fields, at Warwick it is used largely for trace analysis in semiconductor samples. Under ultra-high-vacuum (UHV) conditions, the technique uses a beam of ions (the *primary* ions) accelerated to an energy between 1 and 15 keV ion^{-1}, impacting on the surface of a sample. The impact of each primary ion causes a number of surface atoms to recoil, these *primary recoils* causing in turn secondary and tertiary recoils, etc. The collective name for these motions is the *collision cascade*. The deposited energy and momentum find their way back to the surface of the sample, causing the emission of *secondary particles*: electrons, atoms, and molecules, in varying states of excitation and ionisation. The ionised fraction are diverted electrostatically into a mass spectrometer: quadrupole, magnetic sector (or double focusing), or time-of-flight spectrometers are used, though this thesis will mostly concern quadrupole SIMS. Generally, ions of a particular charge-to-mass ratio are counted over fixed periods of time as continued sputtering causes the erosion of the sample.

Quantification procedures are used to attempt to obtain a *concentration depth profile* from the raw signal: an intensity versus time plot. Due to the many possible forms of atomic motion that can be caused or enhanced by the primary ions, a discrepancy may exist between the quantified depth profile and the *internal* (i.e. 'true') depth profile. In many cases a concentration depth profile of the sample using other techniques is not available (or possible). Here the discrepancy may only be inferred from a

dependence of the depth profile on the analysis conditions (particularly the primary ion energy and angle-of-incidence). Profile distortion is particularly visible on profiles of thin features or large concentration gradients. For semiconductor analysis, the rapid development of typical samples in recent years has made the distortions increasingly obtrusive and troublesome.

1.2. Objectives of this research.

Quantification methods must be designed to invert the effects of the analysis. Current quantification procedures do not yield accurate results because the model used to derive conventional quantification methods is far too simple. The main objective of this research was to take the first step toward a single, complete, self-consistent quantification process. To achieve this a more accurate process model and a suitable method for the inversion was required.

The convolution integral had previously been suggested for the empirical modelling of distortions during SIMS depth profiling (e.g. Werner, 1982, King and Tsong, 1985, and Clegg and Beall, 1989). The first stage of this research was to develop an empirical model for the analysis process, based upon the use of the convolution integral.

The second stage was to assess methods for the inversion of the model. As the solution of a convolution integral is undefined (except in the analytic case) many proposed deconvolution methods are numerically unstable and require smoothing. This research was thus to concentrate on iterative methods for the inversion, particularly Maximum Entropy (MaxEnt) reconstruction. This method draws on information theory and probability calculus to select one of the many possible solutions as best (in the sense of

least biased). It had been used with great success in other fields, notably astronomy and spectroscopy. However, SIMS depth profiles may have a signal to background ratio of 10^6, and are routinely displayed on a logarithmic plot. The large importance placed on such tiny features makes SIMS a unique application of MaxEnt reconstruction.

The third stage of the research was to concentrate on assessing the performance of the method.

1.3. Thesis Overview.

In chapter two we present the simplified model for the sputtering and ionisation processes, and the quantification processes currently applied to dynamic SIMS data. As the quantification is known to be approximate, we shall present the adjustments made to the procedures which yield, under certain circumstances, more accurate concentration profiles. We shall then review previous attempts to remove the distortions from depth profiles using deconvolution. Stressing that the quantification processes should invert the model for the measurement, we detail some of the intrinsic processes that are missing in the simple model, and briefly survey more accurate models.

The MaxEnt method is introduced in chapter three. We first demonstrate that even if the convolution is a good approximation for the analysis process, simple deconvolution methods are inappropriate. We then present the argument for the use of the MaxEnt method for the reconstruction of depth profiles. This method was specifically developed for use on incomplete and noisy experimental data.

In chapter four we develop the (empirical) model for the distortions in a SIMS depth profile, using the convolution integral. We stress, however, the

limitations of this model. We argue that the response function for a SIMS instrument can be measured from suitable samples, and give the formula for a fully quantitative response function. As the MaxEnt method is based on the statistical properties of measured data, an estimate of the noise on the data is required. In this chapter it is attempted to deduce the likely distributions of noise on a SIMS depth profile, given the nature of the experiment.

The iterative nature of MaxEnt requires an algorithm to search for the maximum of a function. This *optimisation* stage may be time-consuming, thus it is important that an efficient algorithm is used. To allow rapid development of the quantification process it is also important that a conceptually simple algorithm is used. In chapter five we describe the design and development of the three optimisation algorithms used in this research. These algorithms have been assessed for their speed in providing the required solution.

Many classes of experiment have been performed during this research. To avoid a 'grainy' feel, the experimental methods have been grouped into chapter six. We first describe the EVA2000 quadrupole SIMS instrument used to obtain the depth profiles, then the method for the characterisation of the noise on a SIMS depth profile. We then introduce the simulated data used for the assessment of the speed of the various algorithms, and that used to assess the performance of the MaxEnt method on various types of features commonly found in SIMS. A number of genuine depth profiles have also been used: these too are described here.

Chapter seven contains the collected results from the various experiments described in chapter six, and discussions of the results and implications.

In chapter eight we describe the in-house depth profile quantification software 'ProWin', developed by the author during this research. This software has been developed not only to allow the extraction of parameters required by the MaxEnt method, but also to simplify the use of conventional quantification, and to provide the ability to quantify the more sophisticated data expected from the new EVA3000 instrument, currently nearing completion at Warwick.

Finally we present our conclusions, and indicate the direction that, in the author's opinion, future development of the technique should follow.

2.0. Quantification methods in SIMS.

The basis of current dynamic SIMS quantification lies in an oversimplified model for the processes of sputter-induced erosion and ionisation. We shall first introduce this model and the appropriate quantification techniques. The conditions under which this quantification is considered appropriate will be discussed, as will the corrections applied to SIMS data obtained under other conditions. Stressing that the quantification processes invert the *model* for the experiment, we shall discuss the most fundamental shortcomings of the model, and propose that an alternative model will result in more accurate quantification.

2.1. A Simple Model for sputter erosion.

To avoid the proliferation of unnecessary constants this model will be described in self-consistent units, not necessarily those most commonly used for the physical parameters.

2.1.1. Ideal erosion.

It is assumed (Hofmann, 1976) that the primary ion flux [ions cm^{-2} s^{-1}] is uniform across the entire analysed sample area, and does not vary with time. Assuming that the surface under bombardment is initially flat and neglecting any induced topography, the sample is assumed to erode on a layer-by-layer basis. In a single *frame* or *cycle* of time Δt [s] at primary ion current I [A] over crater area A [cm^{-2}], a primary ion dose $\Delta\varphi$

$$\Delta\varphi = \frac{I\Delta t}{eA} \qquad [\text{ions cm}^{-2}] \tag{2.1}$$

where e is the charge per primary particle, causes the erosion of the sample. The depth eroded by this dose, Δz [cm], depends on the surface composition

via the *sputter yield*. In this simplified model it is normally assumed that the depth eroded per frame is constant; the depth eroded after *f* frames is

$$z(f) = f.\Delta z \tag{2.2}$$

2.1.2. Ideal ionisation.

Each depth increment Δz corresponds to an eroded volume V [cm^{-3}], given by

$$V = A.\Delta z \tag{2.3}$$

If the average concentration of atoms of mass m [amu] in the volume V is C_m [atoms cm^{-3}], then in time Δt, VC_m atoms of this species are emitted in various states of ionisation or bonding. Given that ions in charge state q are to be detected, ionised with a probability $\alpha_{m,q}$ [ions (sputtered atom of mass m)$^{-1}$], and the product of the SIMS instrument's transmission and detection probabilities is represented as $\Gamma_{m,q}$, then

$$Y_{m,q} = A\Delta z\alpha_{m,q}\Gamma_{m,q}C_m \tag{2.4}$$

ions will be detected during that frame. Separate measurement of $\alpha_{m,q}$ and $\Gamma_{m,q}$ is extremely difficult, but the product

$$\tau_{m,q} = \alpha_{m,q}\Gamma_{m,q} \tag{2.5}$$

is termed the *useful yield* for element of mass *m*. It can be measured from a suitable standard and is used as a figure of merit for an instrument. Therefore, in this simple model the detected signal in each frame is related to the concentration via

$$Y_{m,q} = A\Delta z\tau_{m,q}C_m \tag{2.6}$$

2.2. The quantification methods.

We shall use the term 'raw' depth profile to describe a plot of measured signal intensity Y [ions frame^{-1}] versus frame number, or the

equivalent form of intensity Y [ions s^{-1}] versus time t [s]. The desired graph is a plot of atomic concentration C [atoms cm^{-3}] versus depth z [μm]. The quantification methods attempt to map one graph onto the other. Given the above model for the process, a raw depth profile is quantified by a two-stage process.

2.2.1. Depth quantification.

The depth scale is obtained by determining the erosion rate [cm frame^{-1}], using the number of frames during the analysis f_{final} and the depth of the crater z_{final} to obtain

$$\Delta z = \frac{z_{final}}{f_{final}} \tag{2.7}$$

The parameter z_{final} is measured post-analysis, typically using surface profilometry or interference microscopy. With care, a reproducibility of 4% may be obtained using these methods. Round-robin studies indicate that larger errors are routinely obtained (Dowsett, private communication).

2.2.2. Concentration quantification.

Two methods exists for scaling the secondary ion intensity Y [ions frame^{-1}] to an atomic concentration C [atoms cm^{-3}]. In a commercial instrument, common practice is to measure the signal for mass m with respect to that for a reference mass r, both in the same charge state q. From equation 2.6 one obtains

$$\frac{Y_{m,q}}{Y_{r,q}} = \frac{\tau_{m,q}}{\tau_{r,q}} \frac{C_m}{C_r} \tag{2.8}$$

which may be written

$$\frac{Y_{m,q}}{Y_{r,q}} = \frac{C_m}{C_r} S_{m,r} \tag{2.9}$$

where $S_{m,r}$ is the *relative sensitivity factor* for mass *m* with respect to reference *r*. The reference signal is commonly chosen to be the matrix channel, and values of $S_{m,r}$ are tabulated by measuring known (normally ion-implanted) standards. The accuracy of SIMS quantification using this method is thought to be five to ten percent in concentration (Benninghoven, Rüdenauer, and Werner, 1987), but this is likely to be true only for a carefully set-up magnetic sector instrument.

The above method may be much less reliable than is thought, particularly for quadrupole instruments. At Warwick, a variant of the above method is used. Provided the relationship is linear as in equation 2.6 the reference channel is redundant: absolute sensitivity factors may be used. Equation 2.6 may be re-written

$$Y = \Omega^{-1} C_V, \tag{2.10}$$

where Ω is an unknown scaling constant. If an ion-implanted standard with a known total implanted dose C_A [atoms cm^{-2}],

$$C_A = \int_0^{z_{max}} C_V(z)\, dz \tag{2.11}$$

is profiled and depth quantified to give a signal *Y(z)*, then the constant Ω can be determined using

$$C_A = \int_0^{z_{max}} \Omega Y(z)\, dz, \tag{2.12}$$

thus

$$\Omega = \frac{C_A}{\int_0^{z_{max}} Y(z)\,dz}. \qquad (2.13)$$

As $\Omega = A\Delta z\tau_{m,q}$ it is necessary to ensure that the measurement of the standard and the sample use craters of the same area. This is not an inconvenience; the standard is measured before each unknown sample to obtain an up-to-date quantification factor, accounting automatically for a slow variation of any unmonitored parameters.

2.3. Applicability of the quantification methods.

The uniform primary ion flux is required to avoid the evolution of macrotopography on the crater bottom (McPhail, Dowsett, and Parker, 1986). This uniformity may be obtained in two ways:

(i) With a stationary macrofocus beam (the instrument operating as an *ion microscope*) apertures may be used to limit the primary ion flux to the more uniform central region of the beam, which has a Gaussian profile if correctly focused. Alternatively, apertures may be used to limit the detected secondary ions to those originating in the central region of the crater.

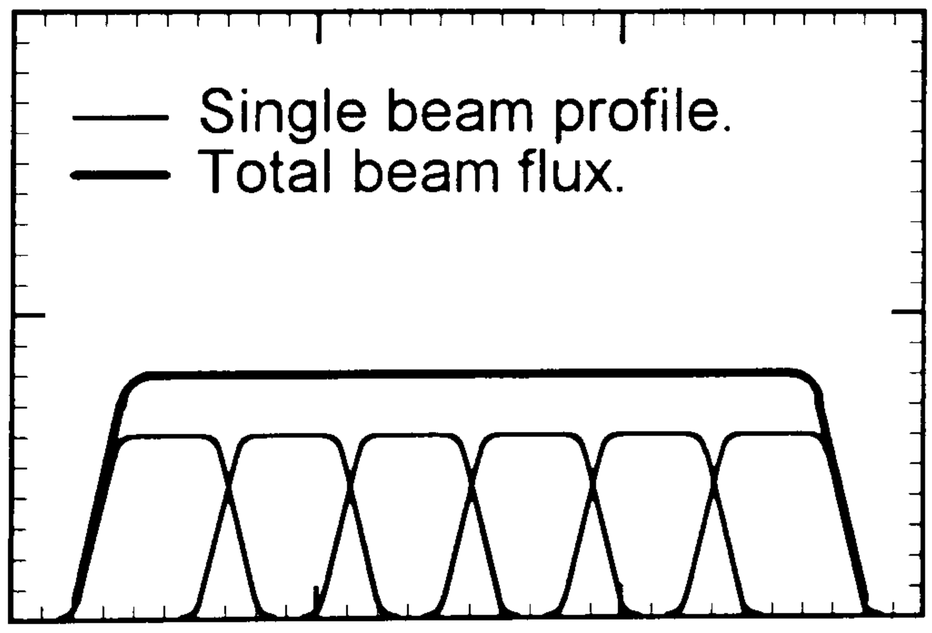


Figure 2.1. Obtaining a uniform ion flux by rastering the beam.

(ii) Increasingly SIMS instruments operate as *ion microprobes*, where a microfocus ion beam is rastered over the crater area (figure 2.1). If this

process is performed accurately, a uniform primary ion flux may be maintained across the crater as required, and the generation of macrotopography may be avoided for at least the first micron (Dowsett, Barlow, Fox, Kubiak, and Collins, 1992a).

The approximation of a constant sputter yield is, in complete generality, poor. Sigmund (1969) demonstrated that the sputter yield depends on the energy deposited at the surface by the primary beam. It is thus very dependent on the near-surface composition. Fortunately, SIMS has the high elemental sensitivity required for trace analysis. At Warwick we specialise in the analysis of semiconductor samples. A common feature of such samples is an elementally pure matrix with very low levels of impurity. During analysis, the surface composition is dominated by the matrix, and the approximation of constant sputter yield is significantly improved. Commonly quoted limits for this 'dilute' effect are a total impurity concentration of a few percent.

The major assumption about the ionisation processes is that the (effective) ionisation probability is constant. The mechanisms and cross-sections for ionisation depend greatly on surface chemistry, and are still poorly understood for most systems. However, for dilute systems the chemistry of the matrix again dominates, and ion yields are approximately constant.

One remaining condition for both sputter yield and ionisation probability is that equilibrium conditions are fulfilled. It is known that, even using noble gas primary ions, the deposition of energy and momentum and the implantation of primary ions cause the development an *altered layer*, with composition and chemical state different from the bulk material.

Conservation of mass implies that, some time after the start of the experiment, the fluxes of all types of particles into and out of the altered layer reach a steady-state equilibrium such that the composition of the altered layer is that of the bulk material (Morgan, deGrefte, Warmoltz, Werner, and Tolle, 1981). Until this steady state is reached, the erosion rate and ionisation probabilities will not be constant. The effects of non-equilibrium conditions are more readily apparent when reactive primary ions are used. For semiconductor samples this is common practice.

In summary, the current SIMS quantification methods are applicable only for dilute impurities after equilibrium is obtained. They are used in other conditions under slightly adjusted forms.

2.4. Quantification of higher concentrations.

If the sample does not fall into the dilute limit the simple model is no longer appropriate and corrective adjustments are made to the quantification methods. There are essentially two factors that could vary in a concentration-dependent manner.

2.4.1. Varying erosion rate.

As depth profiling continues, the changing composition of the near-surface region may cause a change in the sputter yield and thus a change in the depth (and volume) eroded in each frame. This will affect the quantification of both axes; an increase in the sputter yield causes a greater depth to be eroded and creates more secondary particles (although as section 2.4.2 will show, not always an increase in the measured signal).

Adjusted quantification procedures may be used in some circumstances: If the sample under consideration is known to be a multilayer sample, the approximation that the sample is a number of distinct regions with different erosion rates may be better than the (incorrect) linear assumption. Suitable standards may be measured to determine the erosion rate applicable in each region. The transition points are then determined from the raw SIMS signal, and a piece-wise linear depth mapping applied as shown in figure 2.2.

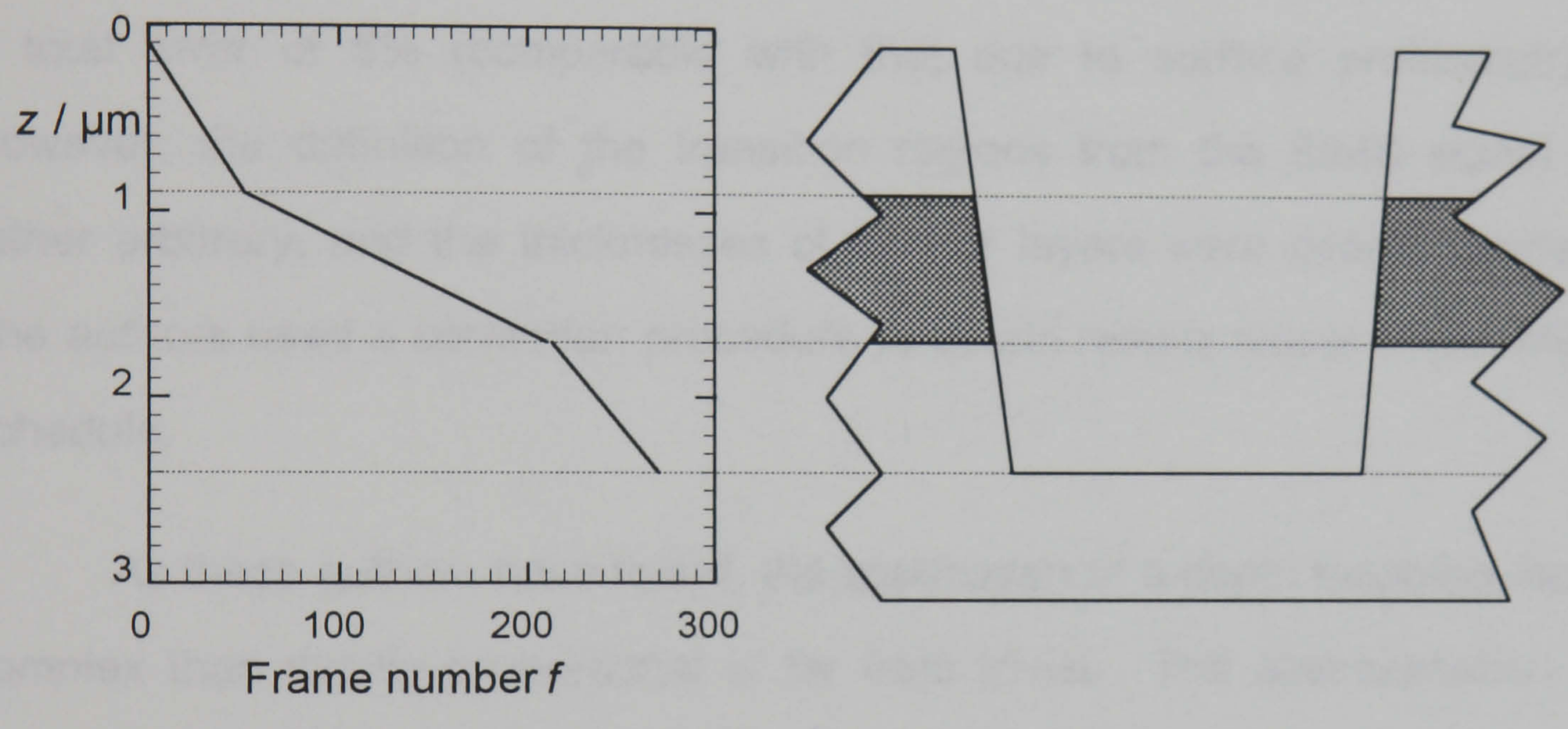


Figure 2.2. A piece-wise linear depth scale.

Voigtmann and Moldenhauer (1988) attempted to determine the erosion rates for multilayer samples using a bevel beside the SIMS crater. The section through the bevel, obtained by surface profilometry, gives directly the depth mapping as in figure 2.2. They claim that as the profilometer scan is continuous across the bevel, the errors are considerably less than the absolute error of the profilometer. The author finds this unlikely as all profilometer measurements are continuous in this sense, implying at least errors of the same magnitude. The error is likely to be larger still: the altered layer on the floor of the bevel traverses regions of different

composition, thus it will not be of constant thickness, and the angular alignment of the bevel scan may cause a subtle change in the bevel profile. Voigtmann and Moldenhauer used marker holes to align the scan, suggested the use of smoothing to reduce the effects of noise on the bevel profile.

A similar depth mapping was performed for AlGaAs/GaAs heterostructures (Meithe et. al, 1991), using sputter yields measured from known standards. A comparison of the calculated crater depth with that due to surface profilometry was used to estimate the errors in the process, giving a total error of 5% (comparable with that due to surface profilometry). However, the definition of the transition regions from the SIMS signal is rather arbitrary, and the thicknesses of thinner layers were over-estimated. The authors used a correction procedure to obtain results closer to the MBE schedule.

As these authors have found, the application of a depth mapping more complex than strictly proportional is far from trivial. The approximation of piece-wise linear erosion is appropriate only if the transition regions are significantly thinner than the layers, and the process involves many more sources of error, as many more measurements are required to determine the mapping. Under other conditions the measurement of a continuously varying erosion rate is required but not yet possible.

2.4.2. Varying ionisation probability.

Variations in erosion rate due to non-dilute conditions are often accompanied by variations of the ionisation probability. These variations strongly modify the simple proportionality of the secondary ion signal to the erosion rate. The effects of these two variations are seldom easy to

separate, but in limited circumstances corrected quantification procedures have been applied.

If the concentration of an impurity exceeds the dilute limit, it may alter its own ionisation probability. Provided the variation is sufficiently slow that equilibrium conditions are maintained (Dowsett and Barlow, in press), the measurement of suitable standards with known composition allows the creation of a calibration curve for the intensity scale. An example is shown in figure 2.3. Each intensity value is mapped onto a concentration value on a point-by-point basis, using interpolation as required. Care must be taken with this technique as the curve is valid only for a particular sample and experimental conditions.

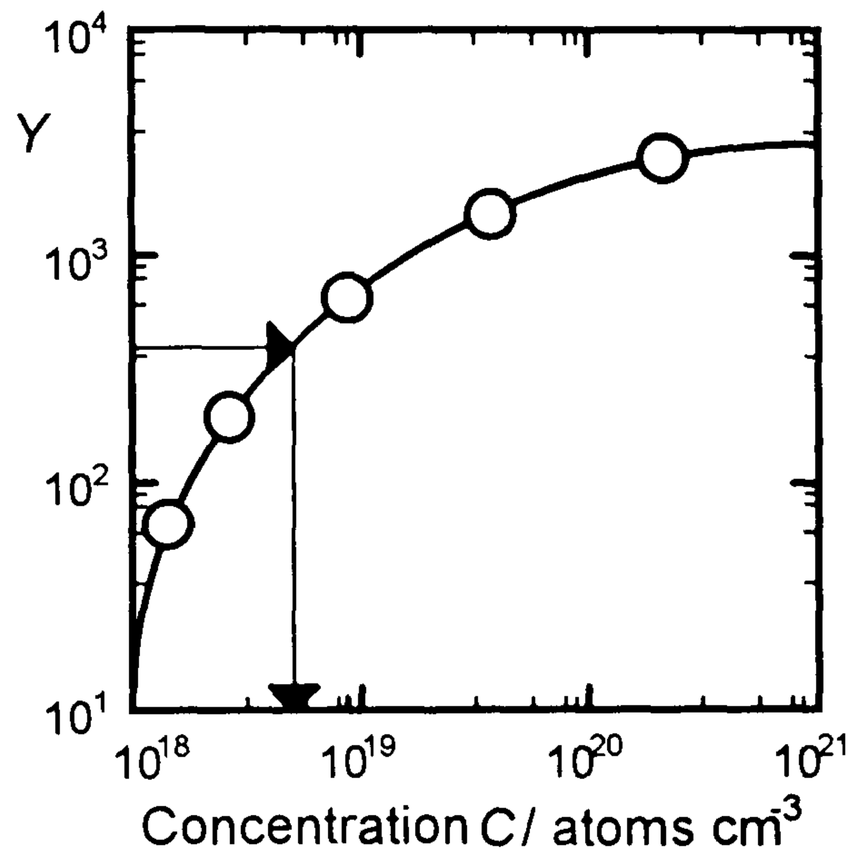


Figure 2.3. A calibration curve for concentration quantification.

For samples where the impurity of interest traverses a number of separate matrices, a piece-wise linear mapping may again be preferable to a linear mapping. A correction for varying ion yields was performed by Spiller and Davis (1985), for samples where it could be demonstrated that no change in erosion rate occurred. Again, quantification in the transition regions is poor.

2.5. The effects of chemically reactive probe ions.

The other major cause of deviations from the simple model is surface composition changes caused directly by the probe ions. To obtain the high sensitivity required for semiconductor applications, reactive probe ions (e.g. O_2^+, Cs^+) are often used to chemically enhance the secondary ion yields. These chemical changes can have the undesired side-effect of adjusting the erosion rate.

One of the best-characterised side-effects is the *differential shift* first noted by Wittmark and Wach (1981). Here, the position of a boron feature in silicon is found to exhibit a dependence on the primary ion energy when using normal incidence O_2^+ ions. The effect was thought to be caused by the change in erosion rate as equilibrium is reached: the pure silicon before equilibrium erodes approximately twice as fast as the silicon dioxide after equilibrium, thus features are shifted toward the surface by an amount dependent on the depth eroded while equilibrium is established. Initial results (on relatively deep craters) suggested that the time-depth relation after equilibrium was still linear and a suitable correction would be to add a compensating shift.

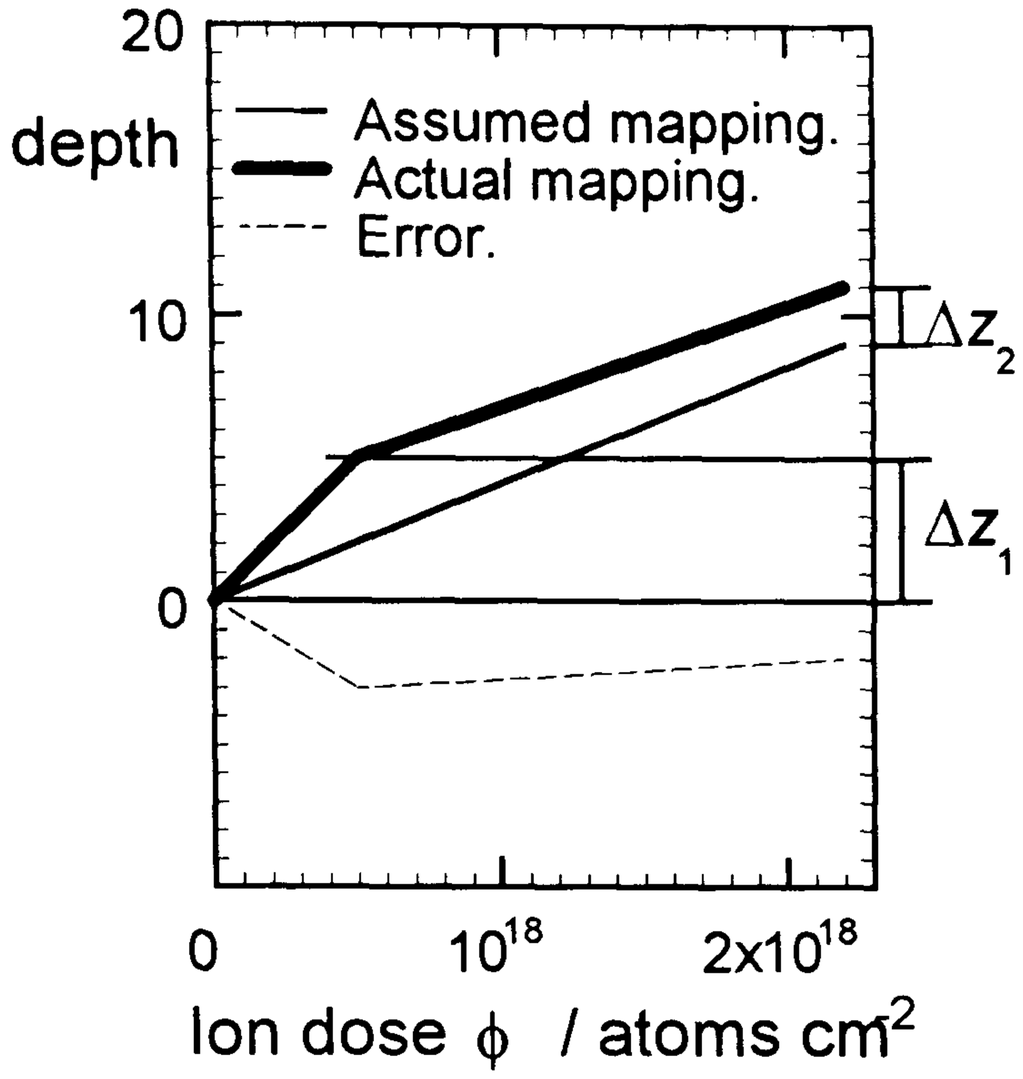


Figure 2.4. Current explanation of the differential shift.

Currently the change in erosion rate is thought to be only one of two effects shifting the profile: the presence of an altered layer on the floor of the crater makes profilometer or interferometer measurement of the crater depth inaccurate. The two effects are shown in figure 2.4, where the pre-equilibrium thickness and altered layer thickness are denoted Δz_1 and Δz_2 respectively. The two effects need not have the same magnitudes: they may cancel to yield the constant error found by Wittmark and Wach, or result in a depth-dependent shift (Barlow, Dowsett, Fox, Kubiak, and Newstead, 1992).

2.6. Deconvolution in sputter profiling.

The distorting effects occurring during a SIMS depth profile are not limited to changes in erosion rate or ionisation probability, even within the dilute limit. The deposition of energy and momentum by the primary particles causes many forms of atomic motion. Most of the processes lead to a broadening of the profile, noted relatively early in the lifetime of the technique (an early review is given in Anderson, 1979). The ensuing loss of 'depth resolution' has lead a number of authors to add further corrections to the

quantification methods. These corrections generally take the form of a deconvolution: it is assumed that the broadening of the depth profile can be modelled using a convolution equation of the form

$$Y(z) = \int_{-\infty}^{+\infty} c(z').r(z-z')dz', \qquad (2.14)$$

where *Y(z)* is the secondary ion signal, *c(z)* the atomic concentration, and *r(z)* the instrumental response. As the convolution integral is common in other fields the form has often been assumed with little or no justification. The response of the instrument is usually normalised; in some treatments all quantities are normalised during the deconvolution stage, to be quantified after the corrections. As the intention is to remove the mass transport effects caused by the primary ion beam, some of the deconvolution attempts in Auger depth profiling are also relevant.

2.6.1. Auger depth profiling.

The first use of deconvolution in Auger depth profiling was reported by Ho and Lewis (1976). Their effort was directed toward removing the effects of surface roughness on the crater bottom. The degradation of depth resolution with depth led them to model the process using a set of response functions $r(z-z',z')$. Two methods were used to determine the (normalised) response function: It was measured directly from a thin layer, and a computational procedure used to remove the effect of the finite thickness. Alternatively it was measured from a step function using differentiation. As the measurement of a response that degrades with depth requires many standards, interpolation was used to provide approximate responses at intermediate depths. The deconvolution was performed using the van Cittert method (van Cittert, 1931) with an automatic stopping criterion giving

approximately five iterations. The authors rediscovered the classic problems of noise amplification, which they reduced by smoothing their data before each iteration.

Hofmann (1980) adjusted the work of Ho and Lewis to cover the blurring effects due to atomic motion in Auger sputter depth profiling. The convolution used was the more common form using a single response (equation 2.14), and the measured signal, was normalised before the deconvolution is performed. The (normalised) response function *r(z)* was assumed to be Gaussian and depth-independent. The standard deviation was obtained from the measured 'depth resolution' Δz, defined as the signal from a step function to drop from 84% to 16% of its peak value. He suggested the method of van Cittert for features without large concentration gradients, and an 'iterative true profile assumption' method (a general reconstruction method) for other cases.

2.6.2. SIMS.

The definition of 'depth resolution' Δz was originally selected because the output of a SIMS depth profile *Y(z)* was assumed to be related to the internal profile *c(z)* via the response function *r(z)* and a cross-correlation integral (Werner, 1982)

$$Y(z) = \int_{-\infty}^{+\infty} c(z').r(z+z')dz' \qquad (2.15)$$

If the response *r(z)* is Gaussian, the measured profile is an error function and measurement of Δz gives immediately (twice) the standard deviation of the response function. The cross-correlation was quickly shown to be equivalent to the convolution integral (Sanz, 1984), and the inverse Fourier

deconvolution method was later suggested to remove the distortions from the measured profile (Benninghoven, Rüdenauer, and Werner, 1987). At this stage there was no mention of possible limits to the applicability of the convolution equation, or of the effects of noise on the measured profile. Although a number of methods were suggested for obtaining the response, no deconvolution results were published.

King and Tsong (1985) attempted deconvolution of SIMS depth profiles using a response function derived from theory. The equations for the atomic transport during analysis were approximated by a diffusion equation, the diffusion coefficient constant over finite depths. The coefficient and depth range were found by fitting measured profiles of dilute titanium markers in silicon. The response function was obtained using the solution of the diffusion equation for the case of an initial delta function (Carter, Collins, and Thompson, 1981). The convolution equation

$$Y(z) = \int_0^\infty c(z')r(z,z')dz' \qquad (2.16)$$

was then simplified using Laplace transforms,

$$Y(s) = \int_0^\infty e^{-sz}Y(z)dz \qquad (2.17)$$

to give

$$Y(s) = C(s)R(s), \qquad (2.18)$$

and the deconvolution was performed using inverse Laplace transforms. Although the method did sharpen test profiles, the appearance of side-lobes forced them to use smoothing; they multiplied *C(s)* by a simple cosine window before the final inverse transform. As we shall show in section 3.1.4, this

smoothing introduces a logical inconsistency, yielding a solution that could not have resulted in the measured profile.

Turner, Keller and Mars (1989) attempted to determine the diffusivity of Be and Si in GaAs using deconvolution. They model the broadening as three separate, independent processes:

1. The leading edge of features was attributed to the finite sampling depth for secondary ions, modelled as an exponential sampling function from the moving surface. The slope was determined from the leading edge of a sharp layer near the surface and considered depth-independent.
2. The trailing edge of features on a profile was attributed to 'knock-on' broadening effects, modelled as an exponential relocation function. The slope was again determined from a near surface sharp layer and considered depth independent.
3. The remaining broadening effects were attributed to crater bottom roughness and modelled as a convolution with a Gaussian function. It was assumed to be depth-dependent and the standard deviation was found by fitting to measured data from a set of layers at increasing depth.

Once the parameters for each process had been determined from a purpose-grown sample, the difference in the standard deviations of the Gaussian parts was used to determine the diffusivity of Be and Si, using the standard sum-of-squares relation for the standard deviations of convolved Gaussian distributions,

$$\sigma^2_{measured} = \sigma^2_{intrinsic} + \sigma^2_{diffusional} \qquad (2.19)$$

and the diffusion coefficient is

$$D = \frac{\sigma^2_{diffusional}}{2t} \qquad (2.20)$$

where t [s] is the diffusion time. It should be noted that equation 2.19 relies on the assumption that both the response function and the diffusion-broadened profile are Gaussian in form. Although the authors recognised the importance of maintaining identical analysis conditions between determination of the instrumental response and measurement of the 'unknown' sample, the method for determination of the instrumental response is very crude, and the physical interpretation of the three parameters is very naive - 'knock-on' effects had long been shown to be insignificant at the primary ion energies used (Littmark and Hofer, 1980), and bombardment-induced broadening is totally neglected. The 'deconvolution' process is merely a correction of one of the parameters, rather than an attempt to produce a more accurate profile.

More enlightened work in SIMS deconvolution has been performed for GaAs samples by Clegg. Using a convolution equation and the approximation that the resolution function is Gaussian in form, diffusion-broadened profiles were corrected for the effects of broadening during analysis using equation 2.19 (Clegg and Beall, 1989). Measured response functions were later used to test the validity of the convolution approximation for the GaAs(Si) system (Clegg and Beall, 1990). These profiles were used to justify a correction curve for growth and decay slopes of sharp layers. The use of a correction curve for slopes alone is flawed, however, as analytic convolution proves that the width of a thin layer also influences the slope (Dowsett and Barlow, in press). In a later publication (Clegg and Gale, 1991), response functions were measured for Si, Be and Al in GaAs, and

used to simulate the profiles expected from rectangular, exponential and Gaussian distributions. Clegg has suggested a trial-and-error deconvolution method suitable for distributions known to be Gaussian, but no results have yet been published.

2.7. Quantification as inversion.

The desired information from a SIMS depth profile is the *internal* distribution of impurities. The role of the quantification processes is to return this distribution, as far as is possible, given the measured data. To perform this task, the quantification processes are designed to invert a model for the relationship between the initial distribution and the measured signal. If the properties of the sample or the analysis conditions are such that the model is inaccurate, the conventional quantification procedures are inappropriate and others must be used.

2.7.1. Shortcomings of the simple model.

The simple model for the erosion and ionisation occurring during a SIMS analysis relies on a large number of assumptions, many of which have always been recognised as approximations. The quantification procedures appropriate to this model are used routinely, partly because no alternative exists. In the past, when typical features were created using ion implantation, the approximations were good and the quantification procedures accurate enough. For features fabricated using the recently developed methods of interrupted MBE growth (and others), the results of a SIMS analysis can give little information to the grower except the integrated impurity concentration. This is because the model for the analysis is incomplete, thus the quantification procedures do not return the internal distribution.

The simple model totally neglects the possibility of 'imperfect' instrumentation. Many mechanisms exist whereby a slight variation in the analysis conditions reduces the reproducibility of the instrument. This is particularly true of the ion guns themselves, which give a current constant only to a few percent over a typical experiment duration. Also, any non-linearity in the rastering electronics may cause the generation of macrotopography on the crater bottom. These cause differences in analysis conditions between the known standard (or the measurement of the relative sensitivity factors) and the measurement of the unknown sample.

Also neglected are the many types of atomic motion that adjust the impurity profile *inside* the sample, before the sputtering takes place. The collision cascade following each ion impact includes both primary relocation and cascade mixing, causing particularly effective mixing in semiconductor samples (Anderson, 1979) although of the two processes cascade mixing has been found to be dominant (Littmark and Hofer, 1980). Although characteristic diffusion in semiconductors is a slower process than in metals, the energy deposition and very efficient damage creation by the primary beam (Wittmark, 1984) may enhance the normal atomic transport, the isotropic atomic motions causing a 'blurring' of features in the profile. Segregation of impurities may also occur during the analysis, particularly at the surface (Zalm, van der Walle, Gravesteijn, and van Gorkum, 1989) or the interface between the altered layer and the undamaged bulk (Wittmark and Menzel, 1987, and Homma and Wittmark, 1990), distorting the profile in unpredictable ways.

In the sputtering itself, neglected processes are the norm rather than the exception. The approximation that all sputtered particles originate in the

topmost atomic layer is good but not true: some 15% of sputtered particles originate at deeper depths (Sigmund, 1989). Microtopography, which *must* exist on the floor of the crater (Sigmund, 1973), allows the detection of particles from a range of depths, and the sputter yield varies strongly with the local angle of incidence. Any variation in the sputtering probability for each of the elements in the sample will cause the composition of the sputtered flux to differ from that of surface.

The ionisation processes are possibly the least well understood of all. Many models for the ionisation mechanisms have been proposed, giving variable agreement with experiment for certain sets of samples. It is quite possible that all effects occur simultaneously, with varying cross sections, in all samples. It is certain that the cross section depends on the surface chemistry. Alterations to the quantification process to compensate for this have been performed only for very small sets of samples.

With regard to improving the model, the effects of imperfect instrumentation are the realm of instrumental development and will not be considered here. Further, it is the job of the experimentalist to determine and select analysis conditions to minimise the intrinsic effects. Nonetheless, there appears at present to be no method to eliminate effects such as atomic transport and surface microtopography. It is the proposition here that it is the role of improved quantification techniques to remove these effects from the profile: the model for the SIMS process should include these effects.

2.7.2. Toward complete models.

Qualitatively, the events following the impact of a primary ion are well understood. Within the framework of the linear (non-interacting) collision cascade model, Sigmund (1969) derived expressions for the sputter yield and

its behaviour as a function of the major impact parameters. A number of computer codes are under development that simulate the sputtering event within the linear cascade regime. Full Monte-Carlo simulations (Anderson, 1987) and Binary Collision Approximation codes (Robinson and Torrens, 1974, and Biersack and Haggmark, 1980) have given further insight into the sputtering process but all are currently too computationally expensive for the modelling of an entire analysis.

Other classes of process model allow simulation of the entire analysis by making greater approximations: Equations including many types of atomic motions were derived using Boltzmann transport theory by Littmark and Hofer (1980). The predicted profiles for thin tracers in silicon under argon bombardment gave such good agreement with experiment that they considered it mostly fortuitous. Modelling the atomic motions as diffusion with effective coefficients over certain regions of the sample, Collins (1978) investigated preferential sputtering and demonstrated that, in the low-concentration regime, some of the non-linear diffusion terms become negligible. This work was extended by King and Tsong (1984), who added a depth dependence to the diffusion coefficient. Solving the equations numerically they obtained improved agreement with experiment: the predicted profiles of a thin (silver in silicon) tracer showing the exponential leading and trailing edges and rounded top characteristic of many SIMS profiles. The diffusion-approximation based computer code IMPETUS (Armour et.al, 1987) has given excellent agreement with measured SIMS data for silicon in gallium arsenide features (Badheka et.al, 1990).

2.7.3. The need for reconstruction.

The different simulation methods and codes use a variety of theories and approximations; one common factor is that the codes are incremental in time: the evolution of the sample under bombardment is tracked using small time or primary ion dose increments. When these models are complete enough that quantification may be designed to invert them, simple algebraic quantification processes will not be sufficient. A more general reconstruction method will be necessary, where a profile is assumed, the analysis simulated, and the assumed profile adjusted according to the agreement with the measured data.

2.7.4. An intermediate model.

The computational complexity of the simulation codes currently used is such that an immediate test of a reconstruction-based quantification method is not feasible: single simulations may take minutes and it is to be expected that a reconstruction will require a large number of simulations to find the solution. As the ion yield from a given surface cannot be calculated the models are also incomplete. For these reasons it was considered useful to develop a reconstruction-based quantification method using a computationally simple, 'intermediate' model. It will turn out (section 4.1) that a suitable model is a convolution integral. The appropriate inversion process will be, at least in part, some form of deconvolution.

3.0. Deconvolution and Maximum Entropy.

In this chapter we shall demonstrate the dangers inherent in many methods of deconvolution, including those described in section 2.6. We shall present the theory underlying the method of Maximum Entropy (MaxEnt). This method uses information theory and probability calculus to invert the deconvolution equation without subjectivity or bias. It is the method to be assessed in this thesis.

3.1. Traditional deconvolution methods.

The convolution integral is common in many areas of physics, and there are numerous methods proposed for the inversion. There is an inherent (and often overlooked) danger in performing a deconvolution. We shall introduce this error with a discussion of the most basic deconvolution method; many other methods are equivalent.

3.1.1. The Fourier transform.

Given some data *F(x)*, where *x* has dimensions of time (or length), we denote the Fourier Transform of *F(x)* as *f(k)*, where *k* has units of time^{-1} (or length^{-1}) respectively. If the data is a continuous function of *x*, the transform may be written as

$$f(k) = \int_{-\infty}^{+\infty} F(x)\, e^{-i2\pi k x} dx \qquad (3.1)$$

The function *f(k)* may also be called the *frequency spectrum* of *F(x)*. Where the transform does exist, the inverse transform

$$F(x) = \int_{-\infty}^{+\infty} f(k) e^{i2\pi k x} dk \qquad (3.2)$$

returns the original function. Other definitions of the transform may be found, especially those where the inverse integration is performed with respect to $\omega = 2\pi k$. In these cases a factor of $\frac{1}{2\pi}$ must be distributed between the forward and backward transforms (Brigham, 1988).

3.1.2. Deconvolution using a Fourier transform.

The property of the transform that concerns us here relates directly to the convolution integral (the proof may be found in Brigham, 1988):

Theorem. For the convolution equation

$$Y(t) = \int C(t)R(t-t')dt' \qquad (3.3)$$

the Fourier transform of y(t) is related to the Fourier transform of c(t) and r(t) by

$$y(k) = c(k).r(k) \qquad (3.4)$$

Thus the convolution equation may in theory be inverted by performing a division in Fourier space to obtain

$$c(k) = \frac{y(k)}{r(k)} \qquad (3.5)$$

and performing an inverse transform.

3.1.3. Incomplete, noisy data.

The above theorem is correct, but in reality the SIMS data are a set of count rates, known over a discrete and limited range of time. An equivalent transform for discrete data does exist: for a discrete function $F(n\Delta x)$ known at N points with constant x-spacing Δx, the transform is defined at the N points $k=0,1,...N-1$, as

$$f\left(\frac{k}{N\Delta x}\right) = \sum_{n=0}^{N-1} F(n\Delta x)\mathrm{e}^{-i2\pi k\, n/N} \qquad (3.6)$$

The inverse transform is

$$F(n\Delta x) = \frac{1}{N}\sum_{k=0}^{n-1} f\left(\frac{k}{N\Delta x}\right) e^{i2\pi k n/N} \qquad (3.7)$$

where the transform pair (3.6 and 3.7) requires both the time- and frequency-domain functions to be periodic with period *N*. For time-limited functions, such as SIMS data, the implied periodicity results in aliasing errors. Provided the sampling time Δx is suitably small, the aliasing errors are small enough that the discrete transform approximates the continuous case, and the convolution property will hold.

The nature of the discrete Fourier transform is such that the convolution equation becomes many-to-one. If a concentration distribution $C_2(z)$ exists which satisfies (or very nearly satisfies)

$$C_2(z) * R(z) = 0, \qquad (3.8)$$

then this distribution may be added to the true distribution without affecting the convolution. The solution of the convolution equation is said to be underdetermined or *ill-posed*, a phrase coined by Tichonov (Tichonov and Arsenin, 1977).

The inevitable random noise on measured data exacerbates this problem (Jones and Misell, 1970). If the measured profile $Y(z)$ contains measurement noise $\sigma(z)$ such that

$$Y(z) = \int C(z)R(z - z')dz' + \sigma(z), \qquad (3.9)$$

then the associative properties of the Fourier transform cause the division to yield

$$c(k) = \frac{y(k)}{r(k)} - \frac{\sigma(k)}{r(k)}. \qquad (3.10)$$

As measurement noise is random, it contains components at all frequencies. The blurring nature of the response causes it to approach zero for high frequencies, thus the second term in the division becomes large, swamping the data with amplified noise.

There are a host of other suggested deconvolution methods, one of the most often suggested being the iterative method of van Cittert (1931). Almost all are equivalent to this division in Fourier space, and all suffer the same problems (Cooper, 1977).

3.1.4. Correcting for the noise.

The potential gains of using deconvolution have lead to a great many attempts to solve the problem of noise amplification. All these methods involve some form of smoothing. A great many of them are directly equivalent to a modification of the response function such that the noise frequencies are not amplified, such that the solution obtained is

$$c'(k) = \frac{y(k)}{w(k)r(k)} \tag{3.11}$$

where *w(k)* is some weighting function, termed a *frequency window*. Attempts to adjust the direct Fourier division in frequency space range from a simple truncation of the spectrum (Stokes' method), to sophisticated smoothing using smoothly varying windows in frequency space (Harris, 1978 or Nuttall, 1981). Multiplication in frequency space is equivalent to convolution in the time domain with the transform of the window, *W(z)*, thus any running average smoothing (performed before or after deconvolution) is again equivalent. The equivalence becomes harder to discern for other methods of smoothing such as fitting polynomials or splines (Yu, Prutton, Larson, Pate, and Poppa,

1982), but the fundamental problem remains: the result of a smoothed deconvolution is not a solution of the initial equation.

$$C'(z) * R(z) \neq Y(z) \tag{3.12}$$

This logical inconsistency is very hard to avoid using deconvolution methods.

3.2. Inference and Probability theory.

The *ill-posed* nature of the deconvolution problem means that a given set of measured data cannot determine a single solution. The data can be used, however, to *infer the likelihood* of any proposed solution. To infer which of a set of propositions is the most likely we require a method of ranking our belief in each of the propositions. We may then select the distribution with the highest ranking as the most likely. To avoid arguing in circles, the method should be both consistent and logical. By theorising that such a ranking should be performed by assigning a positive real number to each proposition, and postulating two simple mathematical operations that the ranking system should obey to remain consistent and logical, Cox (1946) demonstrated that the set of numbers obtained could be mapped to another set of positive numbers which obey the normal rules of probability theory. Thus *any method of logical and consistent reasoning is equivalent to the use of ordinary probability theory, where the probabilities represent our beliefs, or state of knowledge, about the various propositions*. In the case of SIMS quantification, the probability would represent our belief that the true distribution is a particular proposed *C(z)*.

3.2.1. Bayes Theorem.

Given that probability theory is the language we should use, indeed *are* using, one particular theorem is extremely important - Bayes theorem

(Bayes, 1763). The theorem preceded Cox's axioms by 200 years but may be deduced directly from them. It tells us how to adjust our probability in the light of new measured data. Denoting the probability that *A* is true as *P(A)*, the probability that *B* is true given that *A* is true as *P(B|A)*, and the probability that *A* and *B* are both true as *P(A,B)*, the theorem states that

$$P(A,B) = P(B|A)P(A) \tag{3.13}$$

In terms of the current problem, we require *P(C|Y)*: the probability that the true distribution is *C(z)*, given the measured data *Y(t)*. Using Bayes theorem,

$$\begin{aligned} P(C,Y) &= P(C|Y)P(Y) \\ &= P(Y|C)P(C) \end{aligned} \tag{3.14}$$

thus

$$P(C|Y) = \frac{P(Y|C)P(C)}{P(Y)} \tag{3.15}$$

The terms may be given descriptive names (Skilling, 1991): *P(C)* is termed the *prior* probability - the probability we would assign to distribution *C(z)* *before* we make any measurement. It has been the subject of much debate. *P(Y)* is termed the *evidence*, but it can be seen as a normalising constant. *P(C|Y)* is the quantity we require - the *inference*. *P(Y|C)* would be termed the *likelihood* probability - how likely are we to measure *Y(t)* from a distribution *C(z)*. The form of the likelihood probability depends on the nature of the experiment.

The fact that our belief in a distribution *after* the experiment is affected by our belief in the distribution *before* the measurement is not obvious. However, it should be noted that if the data are accurate the likelihood probability will be sharply peaked and the prior information largely irrelevant. If the data are inaccurate, or unrelated to the quantity we require, then our

state of knowledge (or belief) after the experiment will be largely the same as before it. An excellent demonstration of the use and importance of prior probabilities is given by Sivia (1990).

3.2.2. The inclusion of prior information.

An important aspect of the Bayesian probability argument is the ability to include prior information: information available before the experiment is performed. Without a single analysis it is known that negative concentration values are meaningless. This fact represents a considerable amount of prior information - it allows us to rule out a great many possible distributions before the experiment is performed. It is completely ignored by traditional deconvolution methods. Within the Bayesian probabilistic derivation, such information is easily incorporated. Denoting any previous information (or assumptions) as I, equation 3.15 can be re-written as

$$P(C|Y,I) \propto P(Y|C,I)P(C|I) \qquad (3.16)$$

Thus our prior information or assumptions affect the *posterior* probability by affecting both the *prior* and the *likelihood* probabilities.

3.2.3. The likelihood probability.

Given a proposed distribution $C(z)$ we may use our model for the analysis to calculate the expected result of measurement, $Y_{calc}(z)$. Knowledge of the distribution of the noise on the data allows calculation of the probability of obtaining a measured value $Y(z)$ if the true value is $Y_{calc}(z)$. For example, if the data are expected to contain Gaussian distributed noise with zero mean and standard deviation $\sigma(z)$ the probability of measuring (in a single frame) Y counts given a 'true' value of Y_{calc} counts, is (Daniell, 1991)

$$P(Y|Y_{calc}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(Y - Y_{calc})^2}{2\sigma^2} \right) \qquad (3.17)$$

The probability that the entire measured profile matches the 'true' profile is thus proportional to

$$P(Y|C) = \prod_z \exp\left(-\frac{(Y(z) - Y_{calc}(z))^2}{2\sigma^2} \right)$$

$$= \exp\left[-\frac{1}{2}\chi^2 \right] \qquad (3.18)$$

where χ^2 is the standard misfit statistic

$$\chi^2 = \sum_t \frac{(Y(z) - Y_{calc}(z))^2}{\sigma(z)^2} \qquad (3.19)$$

We note in passing that Gaussian noise with zero mean potentially conflicts with the positivity of the SIMS depth profile data. In section 4.4.2.2 we shall demonstrate how this inconsistency may be avoided.

3.2.4. The prior probability - Maximum Entropy.

The prior probability distribution has been the subject of much debate. Objections by orthodox statisticians to Bayesian statistics were based, not on the concept that prior information was irrelevant, but on the fact that Bayes theorem gives no method to determine the prior. One of the first attempts to produce a prior was Laplace's *Principle of Insufficient Reason*, which stated that different events should be assigned the same probability if there is no evidence to the contrary. This theorem, while intuitively correct, is difficult to apply quantitatively and has long been abandoned (Jaynes, 1957). The 'personalistic' school of Bayesian statistics (Savage, 1954) offered the

alternative view that, as all probabilities are by nature personal, each person may determine his own prior. This loss of objectivity in statistical inference generated criticism from both orthodox and Bayesian statisticians, and has been largely discredited (Jaynes, 1968).

Any objective prior must avoid bias, yet agree with whatever information is available. The major advance came from the field of information theory. Shannon (Shannon and Weaver, 1949) demonstrated that a quantity representing the 'degree of uncertainty' in a probability distribution *F(x)* is represented by the formula

$$H(F) = -\int F(x)\log(F(x))dx \qquad (3.20)$$

Due to it's similarity with the thermodynamic property, *H* is termed the *entropy* of the distribution. It is the number of bits of information (or questions with yes/no answers) needed to fully determine the distribution. Jaynes (1957) proposed that to use any prior probability distribution other than that with the maximum entropy consistent with prior *testable* information would be equivalent to some arbitrary assumption of information we do not have. We note, in passing, that as the entropy is undefined for negative values, the use of maximum entropy implicitly includes the prior knowledge that *C(z)* is positive.

Shore (Shore and Johnson, 1980) and Johnson (Johnson and Shore, 1983) placed the principle on more solid ground, by deriving the maximum entropy principle from four simple consistency axioms. These axioms said, fundamentally, that if a problem can be solved in more than one way, the results should be consistent. The principle has since been generalised away from probability distributions to any distribution of positive, additive numbers

(Skilling, 1988a, b). According to Skilling, the probability we should assign to a distribution *C*, given testable prior information *I*, is

$$P(F) = exp(\alpha H) \tag{3.21}$$

where α is an unknown dimensional constant and *H* is the generalised Shannon/Jaynes entropy

$$H(F|M) = \int F(x) - M(x) - F(x)\log\left[\frac{F(x)}{M(x)}\right]dx \tag{3.22}$$

The function *M(x)* is a *measure* or *model* - the value to which *F(x)* will default in the absence of other information. If the model is uniform, corresponding to no prior assumptions about the form of the distribution, and both *F(x)* and *M(x)* are normalised, equation 3.22 reduces to 3.20.

3.2.5. The posterior probability - the inference.

Given the expressions for the likelihood probability and the prior probability, the required inference is

$$P(C|Y, I, \alpha, M) = \exp\left(\alpha H - \tfrac{1}{2}\chi^2\right) \tag{3.23}$$

where we have dropped the *evidence, P(Y)* as simply a normalisation constant. Given this probability distribution for the various concentration distributions, it is possible to generate an entire set of likely distributions *C(z)*. As each is by nature a graph, the representation of more than one solution is an added complication: we thus choose to select the *C(z)* corresponding to the peak of the distribution, providing this can be found.

3.3. The Maximum Entropy method.

The MaxEnt method was pioneered by Gull and Daniell (1978) and Skilling (Skilling and Bryan, 1984). The original work preceded the full

probability theory derivations due to Shore and Skilling: It was based directly on the work of Shannon and Jaynes and some optimisation theory.

Here, we seek to find computationally the distribution *C(z)* with the greatest entropy *H*, subject to the *constraint* that *C(z)* could have resulted in the measured data *Y(z)*. A suitable statistic on which to base our decision is that the mismatch

$$\chi^2 = \sum_t \frac{\left(Y_{calc}(z) - Y(z)\right)^2}{\sigma(z)^2} \tag{3.24}$$

Any *C(z)* for which $\chi^2 \leq N$ would be deemed plausible, thus we select the distribution with the greatest entropy such that $\chi^2 \approx N$. A constrained optimisation problem such as this is difficult to solve: the standard method is to combine the constraint with the function to be maximised, via a Lagrangian 'undetermined multiplier'. Thus we perform an *unconstrained* maximisation of the *Lagrangian function*

$$V = H - \frac{\beta}{2}\left(\chi^2 - N\right) \tag{3.25}$$

for some suitable value of β. If, at the optimum, the constraint is satisfied, $\chi^2 = N$, thus *V=H* and the optimum must also be the optimum of *H*. As we are unconcerned with the magnitude of *V* at the maximum, we may rewrite equation 3.25 without the constant terms,

$$V = \alpha H - \frac{1}{2}\chi^2 \tag{3.26}$$

It can be seen that this expression is identical with that due to rigorous probability theory (equation 3.23) except for the exponential term. As the exponential is monotonic it does not affect the maximisation: the two methods are equivalent.

3.3.1. 'Historic' Maximum Entropy.

The dimensional constant α determines the relative weighting of the entropy and the constraint. For a suitably large value of α the optimisation approaches unconstrained optimisation of *H*, yielding a uniform *C(z)* and a correspondingly large χ^2. For small enough α, we are minimising χ^2 without reference to the entropy, and a small value of χ^2 is obtained. The classic method for determining the correct value of α is by trial and error: we select the value of α such that the constraint $\chi^2 = N$ is satisfied at the maximum of the potential function *V*. This method, known as *historic* maximum entropy, is the one to be investigated here.

3.3.2. 'Classic' Maximum Entropy.

With the rigorous probability arguments that derive equation 3.23 from first principles came an added advantage: it is possible to derive expressions for the probability distribution of α itself (Gull, 1988) and select the most probable. In this way no trial-and-error is required, greatly speeding the optimisation stage. The numerical value of the likelihood attributed to the final solution is far greater than that due to historic MaxEnt (Gull, 1988), but the results are qualitatively the same (Skilling, private communication).

4.0. Theory.

It was demonstrated in chapter 2 that the quantification methods used for SIMS depth profiles are based on the simple model for the sputtering process. We showed that the model omits many of the physical processes and is a good approximation only if the features of interest are relatively broad. We also demonstrated that full, complete modelling of the SIMS process is not currently feasible and suggested that an intermediate model may allow both improvement and development of quantification methods. In this chapter we shall demonstrate that a model including the convolution integral may be used as a quantitative process model, rather than as a normalised correction for atomic mixing. As the MaxEnt method requires an estimate of the noise on a SIMS profile, we shall discuss the potential sources of random and systematic (background) errors and how they can be incorporated.

4.1. The convolution model.

To allow any assessment or development of the MaxEnt reconstruction method we require a model for the SIMS measurement which is sufficiently accurate to be meaningful, yet computationally inexpensive enough to allow development of the method in a reasonable time-scale. In the approximation proposed here we restrict ourselves to the dilute concentration regime. This allows us to consider the two axes to be de-coupled and treat them independently, and to consider the result of measurement as a superposition of the profiles from separate layers. While an unfortunate limitation for a proposed process model, it is no more limited than current quantification procedures.

The two mappings are thus (i) from depth z to primary ion dose ϕ, and (ii) from concentration C to measured intensity Y. Remembering that sufficiently precise measurement of erosion rates is technically quite difficult (sections 2.4 and 2.5), we wish to keep the depth-dose mapping relatively simple. We thus presume that during the analysis the erosion of the sample maps the concentration distribution $C(z)$ onto a distribution 'seen' by the detector as a function of primary ion dose $C(\phi)$, and that the atomic mixing and ionisation processes map this distribution onto the measured $Y(\phi)$. The mappings (as yet undefined) may be written

$$C(\phi) = C(\phi(z)) \tag{4.1}$$

$$Y(\phi) = Y(C(\phi)) \tag{4.2}$$

We now consider the erosion of the sample to proceed on a discrete layer-by-layer basis with no generation of macrotopography, as with the simple erosion model previously described. It was decided that attempts to produce super-resolution were undesirable, thus the discrete layers are chosen to correspond to individual frames rather than atomic planes. The primary ion dose ϕ, although the natural parameter for the erosion, is continuous and thus inconvenient as an index. It is possible to simplify both notation and the transition to computer programming provided that the primary ion current and the crater area are constant. Under these circumstances the primary ion dose is linearly related to frame number f, such that

$$\phi = \frac{I\Delta t}{eA} f \tag{4.3}$$

where I is the primary ion current [A], Δt the frame time interval [s], e the electronic charge [C], and A the crater area [cm^2]. In what follows the frame number f is used to index the erosion.

4.1.1. The mapping of depth to time.

The depth mapping function,

$$f = f(z) \tag{4.4}$$

will depend on the sputter rates of each component of the sample. To a first approximation in the dilute regime, the erosion rate may be considered constant (to within all other measurement errors). The depth mapping is thus approximately

$$f = \dot{z}^{-1} z \tag{4.5}$$

where $\dot{z}$ [nm frame^{-1}] is the erosion rate of the bulk material.

The use of reactive probe ions, especially O_2^+, is standard procedure at Warwick. The differential shift observed when a linear mapping is assumed (section 2.5) cannot be included within this simple linear approximation. There are thus two choices

(i) The differential shift is, at least partly, caused by the chemical effects of the probe. We may make the approximation that there are two distinct regions in the sample: the region initially at the surface which erodes as an elemental target, and the deeper region which reaches the surface as an oxide after equilibrium is established (We shall, for the moment, ignore the existence of a transition region). The depth mapping will thus be a bi-linear mapping (Allen, Dowsett, and Collins, 1993), using two erosion rates:

$$\begin{aligned} f &= \dot{z}_{pre-eq}^{-1} z && z \le z_{pre-eq} \\ f &= \dot{z}_{pre-eq.}^{-1} \cdot z_{pre-eq.} + \dot{z}_{eq.}^{-1} \cdot \left(z - z_{pre-eq.}\right) && z > z_{pre-eq} \end{aligned} \tag{4.6}$$

where $\dot{z}_{pre-eq.}$ is the pre-equilibrium erosion rate, $z_{pre-eq.}$ is the matrix depth of the pre-equilibrium region, and $\dot{z}_{eq.}$ the equilibrium erosion

rate. The measurement of these parameters is far from simple, but has been performed using RBS for the boron-in-silicon system under 4 keV ion bombardment (Dowsett, Jeynes, Clark, Webb, and Newstead, 1990).

(ii) Alternatively, the concentration mapping (which already includes the atomic motions broadening the profile) may be adjusted to include the effects of the differential shift. In this manner the depth mapping is kept as simple as possible (i.e. equation 4.5 is used).

4.1.2. The mapping of concentration to intensity.

The atomic motion processes induced by the bombardment cause the components of each layer to spread themselves between layers. The layers are then sputtered, ionised and detected. In general, both the atomic motions and the ionisation may depend non-linearly on the concentration of the layer and its neighbours. However, for small enough concentrations we expect the broadening effects to become concentration-independent (Collins, 1978). It is also found that, within the dilute limit, the measured signal is linearly proportional to concentration, thus the ionisation probability is constant. The measured signal due to a layer at a depth corresponding to frame f' containing a concentration $C_{f'}$ [atoms cm^{-2}] may thus be written

$$Y_{f'}(f) = C_{f'} . R_{f'}(f) \qquad (4.7)$$

where $R_{f'}(f)$ represents the shape of the layer after mixing, and includes any scaling factors due to the (constant) ionisation probability, secondary ion transmission losses etc. $R(f)$ is termed the response of the instrument to this layer. Each and every layer in the sample will be broadened, ionised, and detected, thus (provided again that the impurity atoms do not interact) the measured signal is the sum of the signals from each layer

$$Y(f) = \sum_{f'} Y_{f'}(f), \text{ or} \tag{4.8}$$

$$Y(f) = \sum_{f'} C_{f'} R_{f'}(f) \tag{4.9}$$

The formation of ion-induced topography on the crater bottom during SIMS depth profiling can cause the depth resolution of the technique to worsen with increasing depth. To include this in the model would require a response function which degrades with depth. Fortunately, under normal incidence O_2^+ bombardment the formation of macrotopography is very uncommon within the first two to three microns, as shown in figure 4.1.

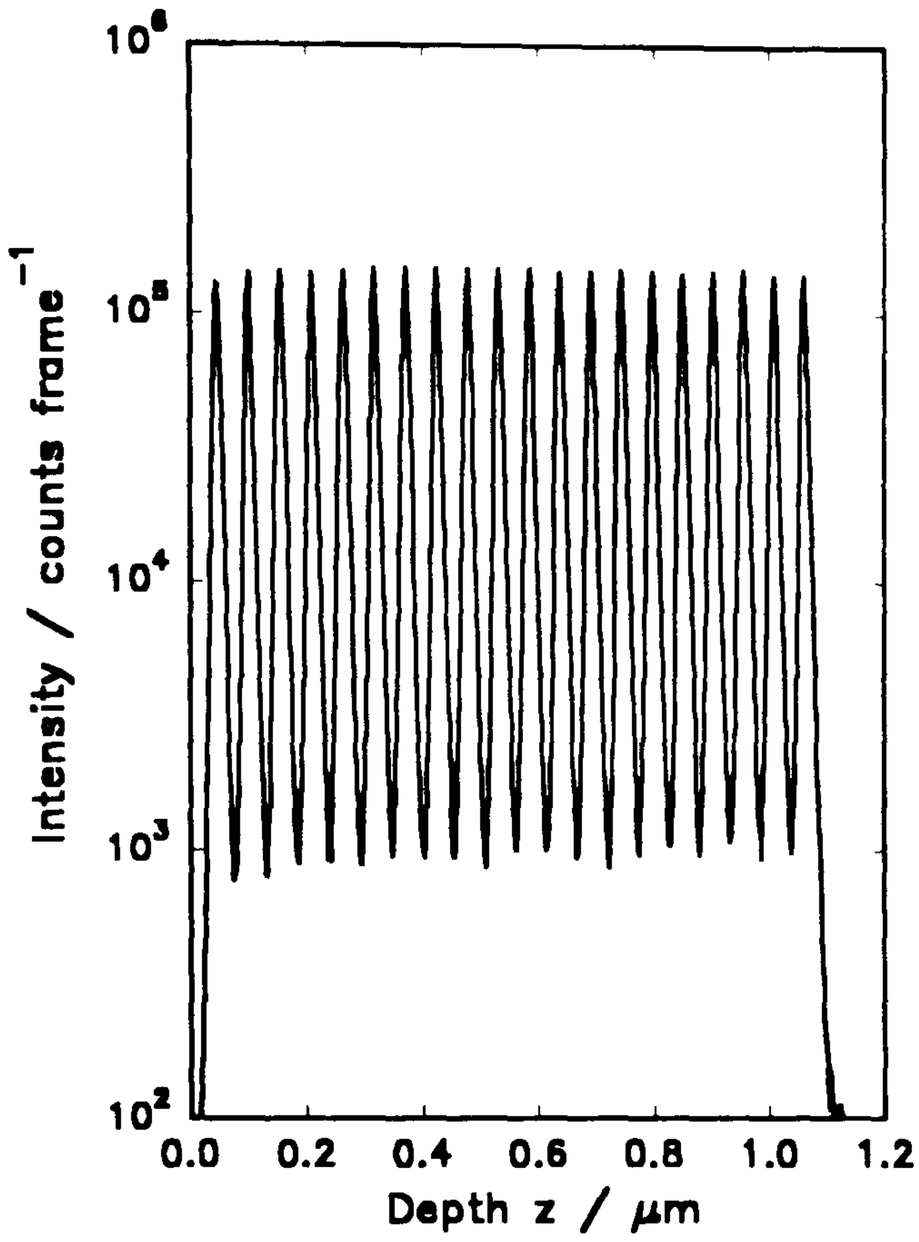


Figure 4.1. Depth profile of SiGe superlattice, demonstrating that the response is depth-independent.

Under these conditions we find experimentally that the profile obtained on measurement of a feature is independent of the depth of the feature, and thus of the frame the layer reaches the surface (Dowsett, Barlow, Fox, Kubiak, and Collins, 1992a). Here, the shape of the response is constant. Assuming a constant differential shift (again a first approximation, in the lack of evidence for a depth-dependence of the shift)

$$R_{f'}(f) = R(f - f') \tag{4.10}$$

and the equation becomes a simple discrete convolution

$$Y(f) = \sum_{f'} C(f') R(f - f') \tag{4.11}$$

This restricted form of equation 4.9 is significantly easier to use, and it requires only one response function to synthesise depth profiles.

The more common form of the convolution is the continuous case, where the summation is replaced by an integration. This can be obtained as follows. Relating the areal concentration C_A [atoms cm^{-2}] to the volume concentration C_V [atoms cm^{-3}], via

$$C_V = \frac{C_A}{\Delta z}, \tag{4.12}$$

where Δz is the thickness eroded during the frame, and writing the discrete convolution now as a function of time,

$$Y(t) = \sum_{t'} C_V(t')\Delta t' \, R(t - t') \tag{4.13}$$

The summation is an approximation of an integral if our layer thickness is small enough, thus

$$Y(t) = \int_0^T C_v(t') \, R(t - t')dt' \tag{4.14}$$

where T is the experimental duration.

4.2. Determination of the response.

For this model to be used to synthesise depth profiles, the response of the instrument must be determined. It is known from the measurement of 'depth resolution' parameters that the profile obtained may be very dependent on the sample and experimental conditions.

4.2.1. Simulation.

Although the sputtering simulation codes are too time-consuming to use as part of an iterative reconstruction method such as MaxEnt, they are

fast enough to use the simulation once only to determine the response of the instrument under the experimental conditions, and proceed using the faster convolution model thereafter. The numerous methods for simulating part or all of the processes occurring during a SIMS depth profile were discussed in section 2.7. Many of these methods give very good qualitative results, or good agreement with experiment if some fitting of physical parameters is allowed. As yet, none of the methods include the ionisation, thus the response would need to be normalised. As this would make the model qualitative, these methods were considered unsuitable.

4.2.2. Measuring the response.

The remaining alternative is to measure the response function. Some authors attempting deconvolution in sputter depth profiling have assumed the form of the response, as Gaussian (Hofmann and Sanz, 1980) or double exponential (Zalm and de Kruif, 1993, also Makarov, 1993), selecting the parameters to match measured profiles. Others have suggested that it may be measured without an assumption about the form from thin layers (if a thinning procedure is used), or from step functions using differentiation (Ho and Lewis, 1976). The author feels that for greatest accuracy the response should be measured *as directly as possible*, with neither assumptions about the form nor fitting procedures.

4.2.2.1. The delta layer.

The response function is the shape that would be obtained on measurement of an ideal delta function of unit areal concentration. Although such features cannot be produced, an extremely good approximation is available. The use of interrupted growth techniques in Molecular Beam

Epitaxy (MBE) and Ion Beam Deposition (IBD) can now produce layers of up to monolayer coverage which are purported to consist of only a single atomic plane of dopant. These so-called 'delta' layers are beyond the depth resolution of any technique in use today, but are known to vary in concentration by orders of magnitude over a few nanometers.

4.2.2.2. The characterisation of delta layers.

The characterisation of delta layer samples is a necessary part of the development of the growth technique. For thickness measurement, SIMS can provide only an upper limit (Mattey et.al, 1990a). Cross-sectional TEM (XTEM) may be used to determine the width if the lattice mismatch is great enough to give strain contrast. As the strain is not necessarily localised to the layer, these measurements are also upper bounds (Mattey et.al, 1990b). Greater accuracy can be obtained in these measurements using X-ray diffraction (XRD), but simple data interpretation is not possible: complex data analysis techniques are required to resolve thin layers (Powell, Kubiak, Whall, and Bowen, 1990). The technique of Z-contrast STEM imaging (Pennycook and Jesson, 1990) can currently resolve individual atom pairs in crystalline silicon, and has been extremely useful in the study of segregation during silicon-germanium epitaxy (Jesson, Pennycook, Baribeau, and Houghton, 1992). It may, in the future, give the capability of direct imaging of delta-doped layers with atomic resolution.

The areal concentration of a delta layer may also be determined by XRD, to an accuracy of about 10% (Powell et.al, 1991a). However, the most accurate technique for the mass analysis of features in the dilute regime is SIMS itself (Powell et.al, 1991b). The areal concentration is, in fact, the only parameter of a delta function which can currently be measured by SIMS.

4.2.2.3. Approximation to the ideal delta function.

Suppose that a proposed delta layer has been grown, and the width has been demonstrated to be sub-nanometer by the characterisation processes. A true delta function has not been grown: the infinite concentration gradients and zero thickness are impossible. Indeed, the MBE process specifically selects stepped wafers to aid the growth process, thus the dopant is unlikely to be contained in a single monolayer across the entire wafer (although it may be only a monolayer thick at any point). Furthermore, the possibility of mass transport during growth must not be neglected. Early attempts to produce delta-doped structures using MBE suffered greatly due to surface segregation during growth. The use of lower temperatures has reduced such problems to beyond the capabilities of the characterisation processes, but it is expected that at finite temperatures some mass transport will occur during growth.

A reasonable quantity of evidence exists, however, that current MBE delta layers are a close enough approximation for the purposes of this research, at least for the boron in silicon system. Firstly, the variation of growth parameters known to affect the thickness of the layer (e.g. growth temperature) below threshold values does not affect the SIMS depth profile obtained on measurement (Dowsett, Barlow, Fox, Kubiak, and Collins, 1992a). Secondly, attempts to produce samples with layers of decreasing thickness result in SIMS depth profiles with a constant shape, but decreasing amplitude. Thirdly, all the depth resolution parameters measured from a (suitably low growth temperature) delta layer vary as a function of the SIMS analysis parameters known to affect the mass transport processes (e.g. primary ion energy and angle of incidence) (Dowsett, Barlow, and Allen,

1994). Each of these effects is an indication that the profile obtained on measurement of a delta layer is entirely due to the SIMS measurement process - a response function. The only evidence to the contrary is that the fitting of a form to the profile of a delta layer results in one parameter that is almost energy-independent (Dowsett, Rowlands, Allen, and Barlow, in press).

4.2.2.4. Availability of appropriate samples.

Delta layers have been produced and characterised at Warwick using MBE for a number of systems. The boron in silicon system has been rather well characterised: boron in silicon delta layers have been grown and characterised as less than 2 nm thick by XTEM (Powell, Kubiak, Whall, and Bowen, 1990), and 0.3 ± 0.5 nm by XRD (Powell et.al, 1991a). An antimony in silicon delta layer has also been grown and was characterised as 1 ± 0.5 nm thick by XTEM (Powell et.al, 1991b).

4.2.3. A Fully Quantitative Definition of the response.

We desire a quantitative model, thus the response must contain the appropriate ionisation coefficient, secondary ion optics transmission and detection efficiency. To this end the response is defined as follows. Assuming that an ideal delta of areal concentration C_A is available

$$C(z) = C_A \qquad z = z_\delta \qquad (4.15)$$

$$C(z) = 0 \qquad z < z_\delta, z > z_\delta$$

Given that the erosion during SIMS analysis maps[1] the depth z_δ onto a frame f_δ the convolution integral predicts that the SIMS profile obtained on measurement of this sample is

---

[1]Using either equation 4.5 (strictly proportional mapping) or equation 4.6 (bi-linear mapping).

$$Y(f) = C_A R(f - f_\delta) \tag{4.16}$$

thus the response could be determined from the measurement using

$$R(f) = \frac{Y(f + f_\delta)}{C_A}. \tag{4.17}$$

The process is shown schematically in figure 4.2.

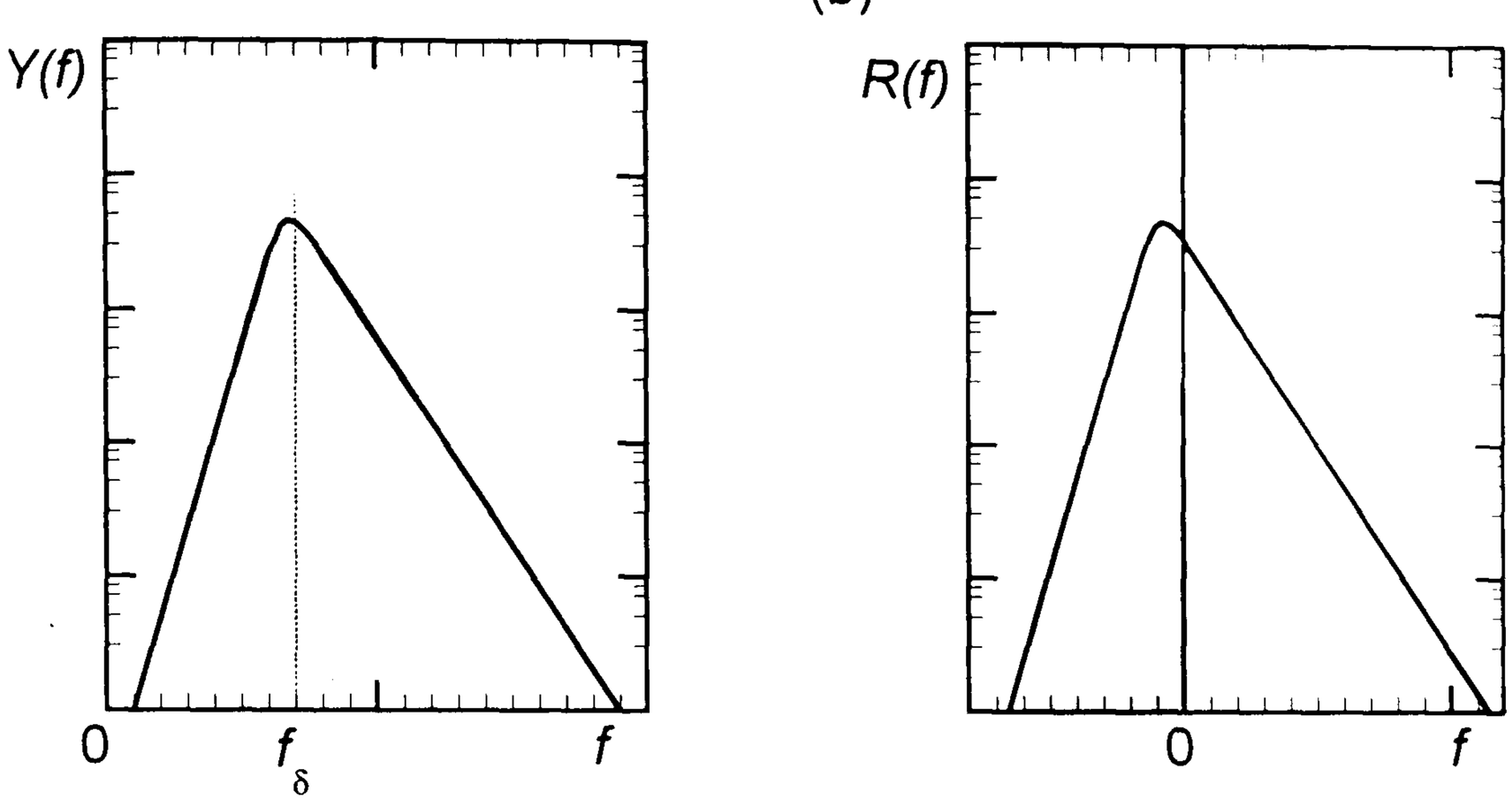


Figure 4.2. Schematic representation of the definition of the response. (a) the measured profile, and (b) the instrumental response.

4.2.4. Calculation of f_δ.

4.2.4.1. Single erosion rate.

If the depth mapping used is the simple linear one, the parameter f_δ [frames] is related to the (true) depth of the delta z_δ [nm] via the erosion rate $\dot{z}$ [nm frame^{-1}]. Given that the depth of the feature as measured by e.g. XRD or XTEM is z_δ, f_δ may be calculated using

$$f_\delta = \dot{z}^{-1} z_\delta \tag{4.18}$$

Using this mapping method, it should be remembered that the response function will implicitly include the differential shift, as due to changing erosion rate, atomic motions, and faulty depth measurement due to swelling on the crater bottom.

If the differential shift (of the centroid of a feature) is known to be σ [nm keV^{-1}] for this impurity/matrix combination, and measurement at energy E_p [keV] predicts a centroid at z_c [nm], we may calculate f_δ *without knowledge of* z_δ, using

$$f_\delta = \left[z_c - E_p.\sigma(E_p)\right]\dot{z}^{-1} \qquad (4.19)$$

Here, knowledge of the true depth of the delta layer has been replaced by knowledge of the differential shift. It can be seen that in this case the response is placed such that the centroid is at $\frac{E_p.\sigma}{\dot{z}}$ (figure 4.3a).

For some systems, where features are symmetric enough that the centroid and peak are at very nearly the same depth, the differential shift is more commonly measured for the peak. Although the author feels that this is a much less reliable measurement of the shift (depending on only a single frame), it provides a very convenient method for placing the response function.

In cases where the true depth of the delta layer is not available, and the differential shift is not well characterised, the least biased assumption that may be made is that no net shift occurs, other than that due to the asymmetry of the response function. In these cases the response may be placed about it's centroid.

4.2.4.2. Bi-linear mapping.

If the mapping used is the bi-linear approximation, two erosion rates $\dot{z}_{pre-eq.}$, and $\dot{z}_{eq.}$, and $z_{pre-eq.}$, and the depth *of matrix* eroded before equilibrium is reached are required. The parameters required can be measured using RBS (Dowsett, Jeynes, Clark, Webb, and Newstead, 1990) although the

experiment has currently only been performed for normal incidence O_2^+ primary ions at 4 keV. A larger range of such measurements is expected from Vandervorst et. al. (Vandervorst, private communication). The parameter f_δ is then

$$f_\delta = \dot{z}^{-1}_{pre-eq.} \cdot z_{pre-eq.} + \dot{z}^{-1}_{eq.} \cdot \left(z_\delta - z_{pre-eq.} \right). \qquad (4.20)$$

This calculation will account for any shift due to the change of erosion rate and the response function will be distributed more symmetrically about the origin, as in figure 4.3b.

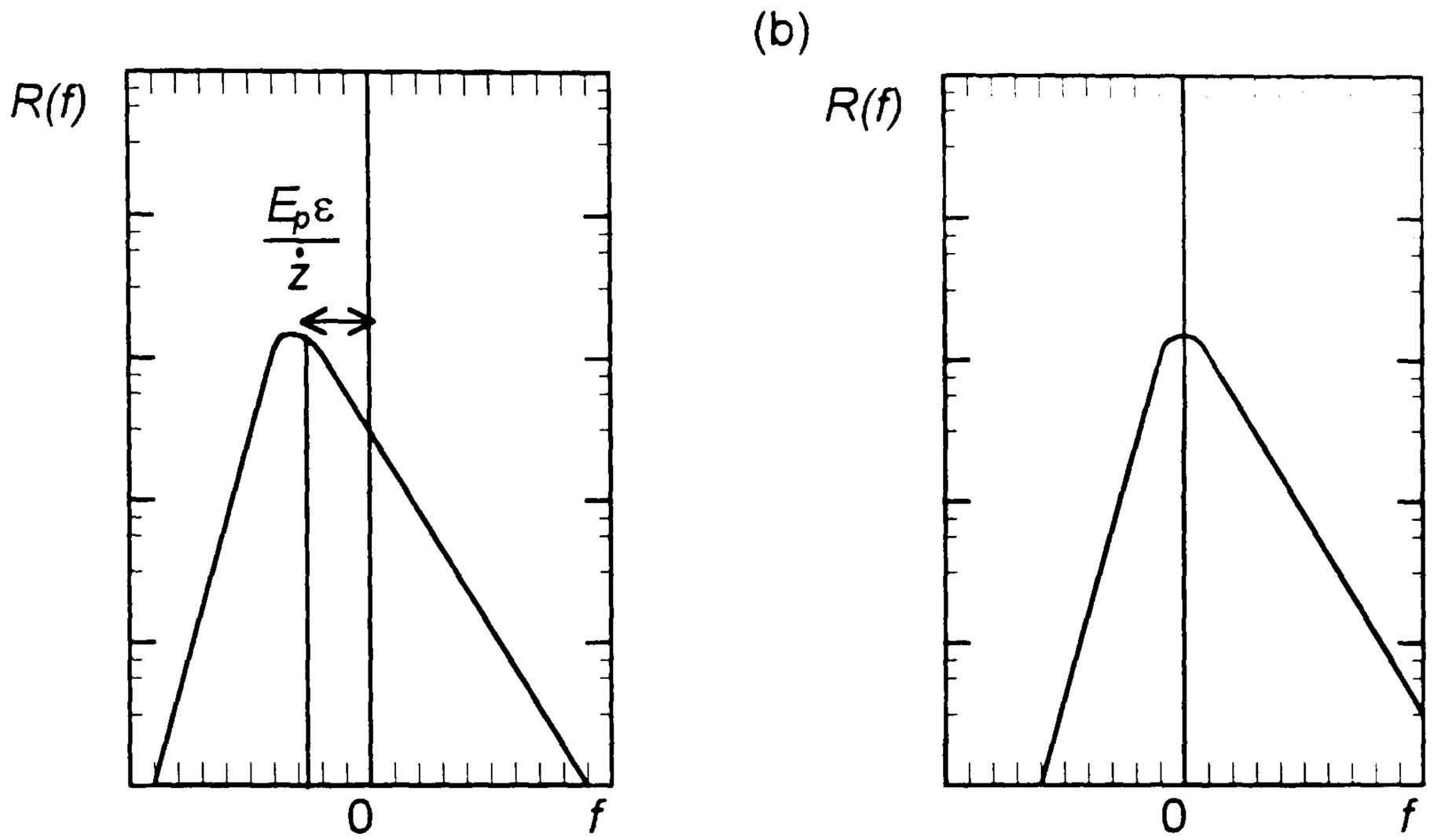


Figure 4.3. Schematic diagram for the response function using the (a) linear and (b) bi-linear depth mapping.

4.3. Determining the erosion rate from a delta layer.

If the depth of the delta layer is known, a significant improvement in the simplicity of the quantification process can be made, within the approximation of a single erosion rate and constant differential shift - the erosion rate can be calculated from the delta layer. In this way the quantification process does not require the measurement of any craters. The

erosion rate can be determined from knowledge of the differential shift as follows:

$$\dot{z} = \frac{z_c + E_p \varepsilon}{f_c} \tag{4.21}$$

where z_c is the depth [nm] at which the centroid of the delta layer occurred, ε is the expected differential shift of the centroid [nm keV^{-1}] at primary ion energy E_p [keV], and f_c is the frame at which the centroid of the delta occurred during the profile.

4.4. The noise on a SIMS depth profile.

In the MaxEnt method the information gained from an experiment is represented (section 3.2) by the likelihood probability,

$$P(Y|C) = \exp\left[-\frac{1}{2}\chi^2\right] \tag{4.22}$$

where

$$\chi^2 = \sum_f \frac{\left(Y(f) - Y_{calc}(f)\right)^2}{\sigma(f)^2} \tag{4.23}$$

The calculation of the likelihood probability requires not only a quantitative model for the measurement process (section 4.1), but also an estimate of the noise on each data point, in terms of the standard deviation σ. The derivation also assumes that the noise is Gaussian distributed with zero mean. Before we may use the MaxEnt method, we must test this hypothesis and determine σ or a method by which to calculate σ from a given depth profile.

4.4.1. Potential sources of noise on a profile.

4.4.1.1. Background signal.

Although mass spectrometry-based techniques have no inherent background signal, the possibility of memory effects in an instrument exists. The ion fluxes in a sputtering experiment can leave material on the walls and other surfaces in the instrument, to be sputtered into the vacuum by a later analysis. Material from a previous analysis may also simply remain in the vacuum. It is likely that these particles will not have the correct velocity and energy distribution to pass through the spectrometer to the detector. Nonetheless, a background signal may exist in some profiles.

4.4.1.2. Random noise.

The distribution of energy and direction of the ions in the primary beam is very small (of the order of 0.1% in energy and 1% in angle). However, this gives rise to a very large variation of impact parameters, thus the secondary ions have a large statistical distribution of energy and velocity. If it is assumed that, for each frame during the analysis, the (unknown) 'true' intensity value is that which would be obtained if every distribution yields it's most probable (mode) value, the actual measured intensity will seldom agree. This will appear as random noise on the depth profile.

There are other effects which can give the appearance of random noise, such as high frequency oscillations in the plasma of the primary ion gun, mains frequency ripple on power supply lines, or pick-up from the radio-frequency quadrupole rods. These signals will be sampled at the frequency of measurement, the frequency- and phase-differences giving a random appearance.

4.4.2. Likely distributions of the noise.

4.4.2.1. Background signal.

The background signal will add a number of counts to each frame. The number of added counts must be a non-negative integer. We may assume that each added count is independent of the others, and if the instrument is operating normally the added counts will occur at a fairly constant rate. A very likely distribution for these counts is thus the Poisson distribution (Eadie et.al, 1971) characterised by a single parameter *m*, which is the mode of the distribution and also the variance. The probability of *x* counts being added to a frame would thus be

$$P(x|m) = \frac{m^x e^{-m}}{x!} \tag{4.24}$$

4.4.2.2. Random noise.

The major cause of random noise will be due to the counting nature of the experiment, thus we expect noise of a Poisson distribution as above. However, as typical signal intensities are in the thousands of counts per frame the Gaussian distribution, characterised by a mean *m* and standard deviation σ

$$P(x|m,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} \tag{4.25}$$

will be an extremely good approximation. (It is also less prone to numerical overflow for large *x*). The standard deviation of a Poisson distribution is related to it's parameter by (Eadie et.al, 1971)

$$\sigma = \sqrt{m} \tag{4.26}$$

thus we may approximate this random noise as a Gaussian with zero mean, provided that the standard deviation is allowed to vary during the profile. For frames with a small 'true' (<8) value the two distributions do differ.

4.4.2.3. Calculating the standard deviation.

From the above discussion it is apparent that it will not be possible to calculate a single σ applicable to every profile. Indeed the standard deviation of the noise in a frame is expected to depend on the number of counts occurring in that frame.

It was thus decided, rather than attempting to characterise two distributions, to use the approximation that the noise in each frame is Gaussian distributed with zero mean. By allowing the standard deviation to vary between frames enough flexibility is retained to estimate the noise across the entire profile. The difference between Gaussian and Poisson for small signal intensities would have two effects

(i) A slight error in the calculation of the mismatch in the background region. As this is seldom the region of interest, this is not a major problem.

(ii) During the MaxEnt search a negative (calculated) signal intensity would be assigned a finite (non-zero) probability. This is also not a problem, as a non-negative concentration distribution cannot yield a negative signal intensity.

We thus assume that the measured profile is the convolution of the true profile with the instrumental response, plus two other terms:

$$Y(f) = \int C(f')R(f - f')\,df' + \overline{Y}_b + Y_r(f) \qquad (4.27)$$

where $\overline{Y}_b$ is the mean number of background counts added to each frame and $Y_r(f)$ is the number of random noise counts added. $Y_r(f)$ is assumed to be Gaussian distributed with zero mean, and standard deviation calculated from the measured profile itself. Some characterisation of the noise was required to find an effective value of $\overline{Y}_b$ and the standard deviation to use.

In the MaxEnt reconstruction the convolution integral will be used to calculate the profile expected from a trial concentration distribution. This calculation must now include the addition of $\overline{Y}_b$ counts to every frame, if $\overline{Y}_b$ is non-zero. This complication (and added computation) may be removed by noting that the chi-squared misfit statistic is identical whether we use

$$\chi^2 = \sum_f \frac{\left(Y(f) - \left(Y_{calc}(f) + \overline{Y}_b\right)\right)^2}{\sigma^2}, \text{ or} \qquad (4.28a)$$

$$\chi^2 = \sum_f \frac{\left(\left(Y(f) - \overline{Y}_b\right) - Y_{calc}(f)\right)^2}{\sigma^2} \qquad (4.28b)$$

thus the background term may be subtracted from the measured data *before* the reconstruction is performed.

5.0. Description of the optimisation algorithms.

A number of algorithms have been used during this research. Each was intended to be an improvement over the last in terms of speed, which is always the major problem in optimisation. The design and development of these algorithms will be described, and flowcharts (but no code) will be given. The measurement of the relative efficiency of each algorithm was done using methods described in chapter 6, with results in chapter 7.

5.1. The initial algorithm.

The first algorithm used was a continuation of work begun by Collins (Collins and Wragg, 1977). Implicit in the theory was the assumption that the depth-frame mapping was strictly linear, thus we may speak in terms of z directly. The problem was transformed to give extra information about the unknown concentration distribution $C(z)$. We shall describe the stages leading from $C(z)$ to this function, then describe the search algorithm used.

5.1.1. Representation of the distribution.

5.1.1.1. As a Histogram.

The speed of an optimisation algorithm depends primarily on the number of variables. Accordingly the concentration distribution $C(z)$ was represented as $w(x)$: a general histogram with N blocks. Each block may be represented by two co-ordinates, w_j and x_j, $j=1..N$. To simplify the algebra Collins assumed the existence of an ordinate at $(w_0,x_0)=(0,0)$. It was hoped that this representation would allow detail where it was needed, but keep the number of ordinates small, as large areas with no detail could be represented by a single broad block. The number of blocks was to start at three, where

the optimisation would proceed most quickly, and be increased during execution until some desired resolution was reached.

5.1.1.2. Normalisation.

The algorithm made great use of normalisation. Both the 'depth' x and 'concentration' w were normalised such that

$$x = \frac{z}{z_{max}} \tag{5.1}$$

$$w(x) = \frac{z_{max}}{C_T} C(x), \tag{5.2}$$

where z_{max} is the maximum depth eroded during measurement and C_T is the total (integrated) concentration

$$C_T = \int_0^{z_{max}} C(z)\,dz. \tag{5.3}$$

Accordingly the normalised concentration distribution $w(x)$ obeys

$$\int_0^1 w(x)\,dx = 1 \tag{5.4}$$

5.1.1.3. Transformation of variables.

Some transformations of variables were performed to simplify the calculation of the entropy. The normalisation of the distribution was the first stage, simplifying the formula for the Shannon-Jaynes entropy (equation 3.20) to

$$H(w|m) = -\int_0^1 w(x) \log\left(\frac{w(x)}{m(x)}\right) dx \tag{5.5}$$

where *m(x)* is the measure or model (section 3.2). Initially Collins assumed that the model would be maximally non-committal, to 'allow the data to speak for themselves', thus

$$m(x) = 1, \qquad 0 \le x \le 1. \tag{5.6}$$

In this case equation 5.5 becomes

$$H(w) = -\int_0^1 w(x)\log(w(x))\,dx \tag{5.7}$$

To further simplify equation 5.7, Collins introduced the cumulative concentration distribution *W(x)*, such that

$$W(x) = \int_0^x w(x')\,dx' \tag{5.8}$$

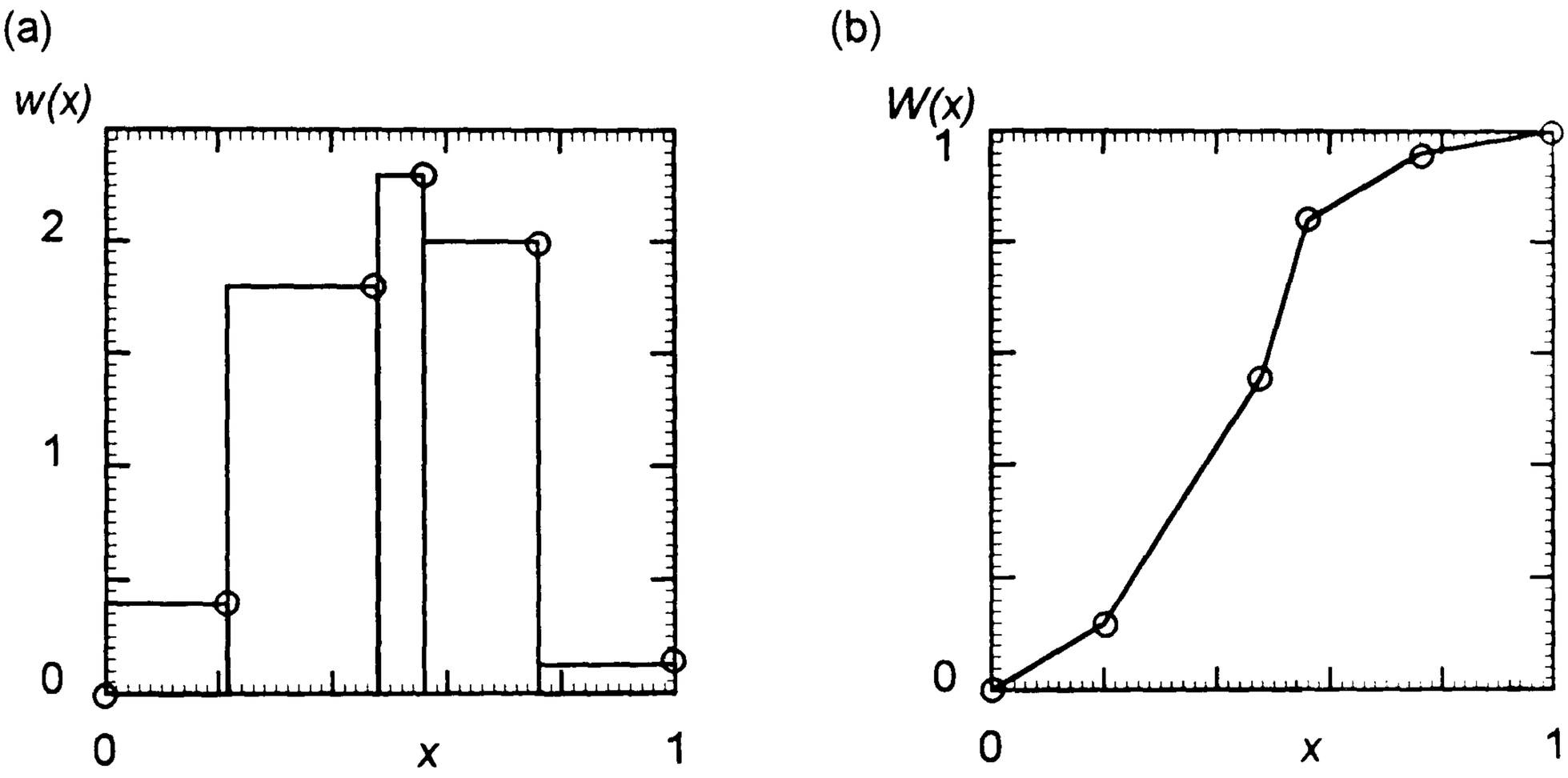


Figure 5.1. The representation of the concentration distribution as (a) a histogram w(x), and (b) a cumulative polygonal arc.

W(x) thus satisfies

$$W(0) = 0, \text{ and} \tag{5.9}$$

$$W(1) = 1$$

and *must* be a monotonic function of *x* (this was considered one of the most important pieces of extra information). Joining the ordinates with lines, rather

than as a histogram, he obtained the *polygonal arc*, as shown in figure 5.1b. He then inverted this relationship and attempted to search for the function *x(W)* where we now know that the unknown function must satisfy

$$\begin{aligned} x'(W) &\geq 0 \\ x(0) &= 0 \\ x(1) &= 1 \end{aligned} \qquad (5.10)$$

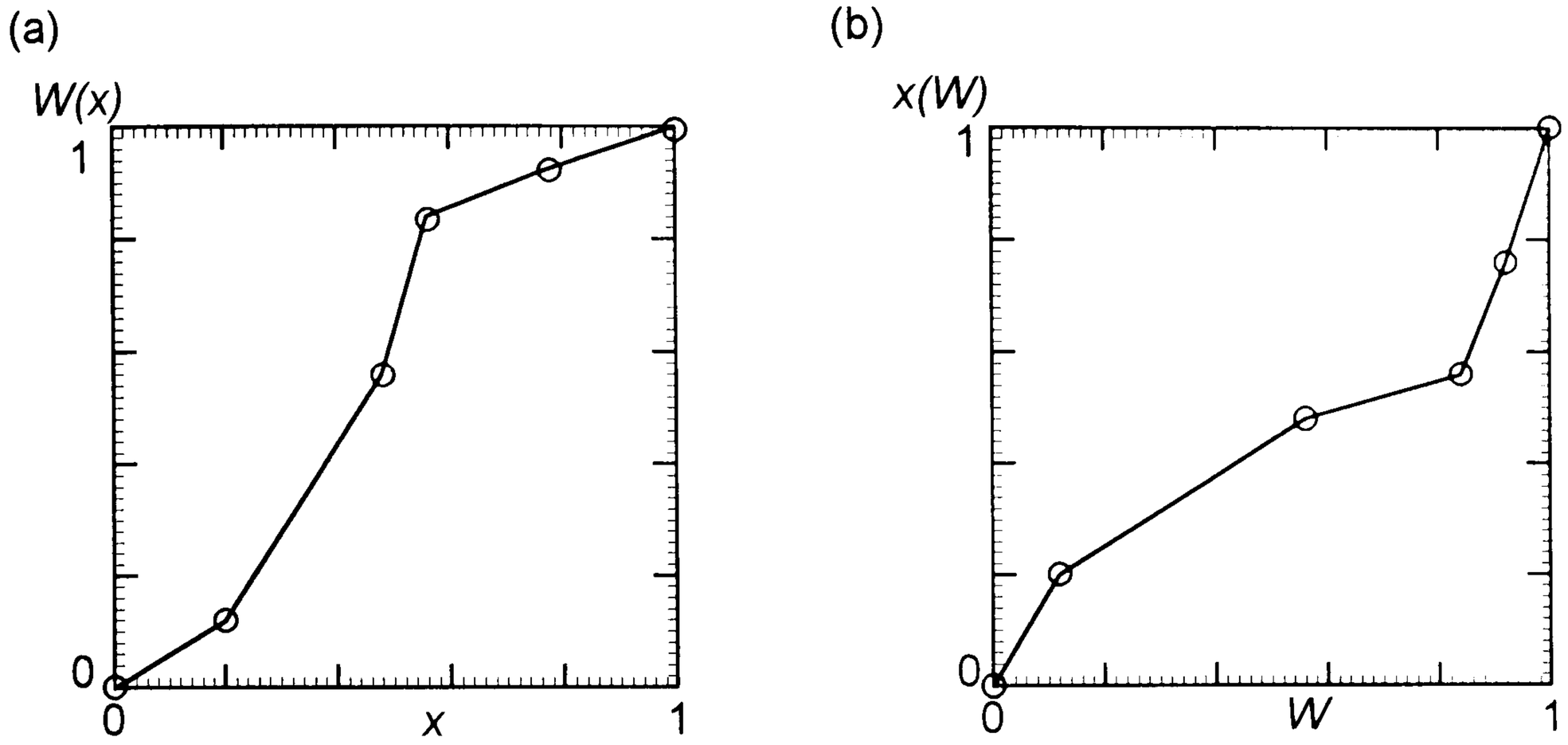


Figure 5.2. The representation of the distribution as (a) a polygonal arc W(x) and (b) the corresponding x(W).

The formula for the entropy (equation 5.7) is now

$$H(x(W)) = \int_0^1 \log(x'(W))\,dW. \qquad (5.11)$$

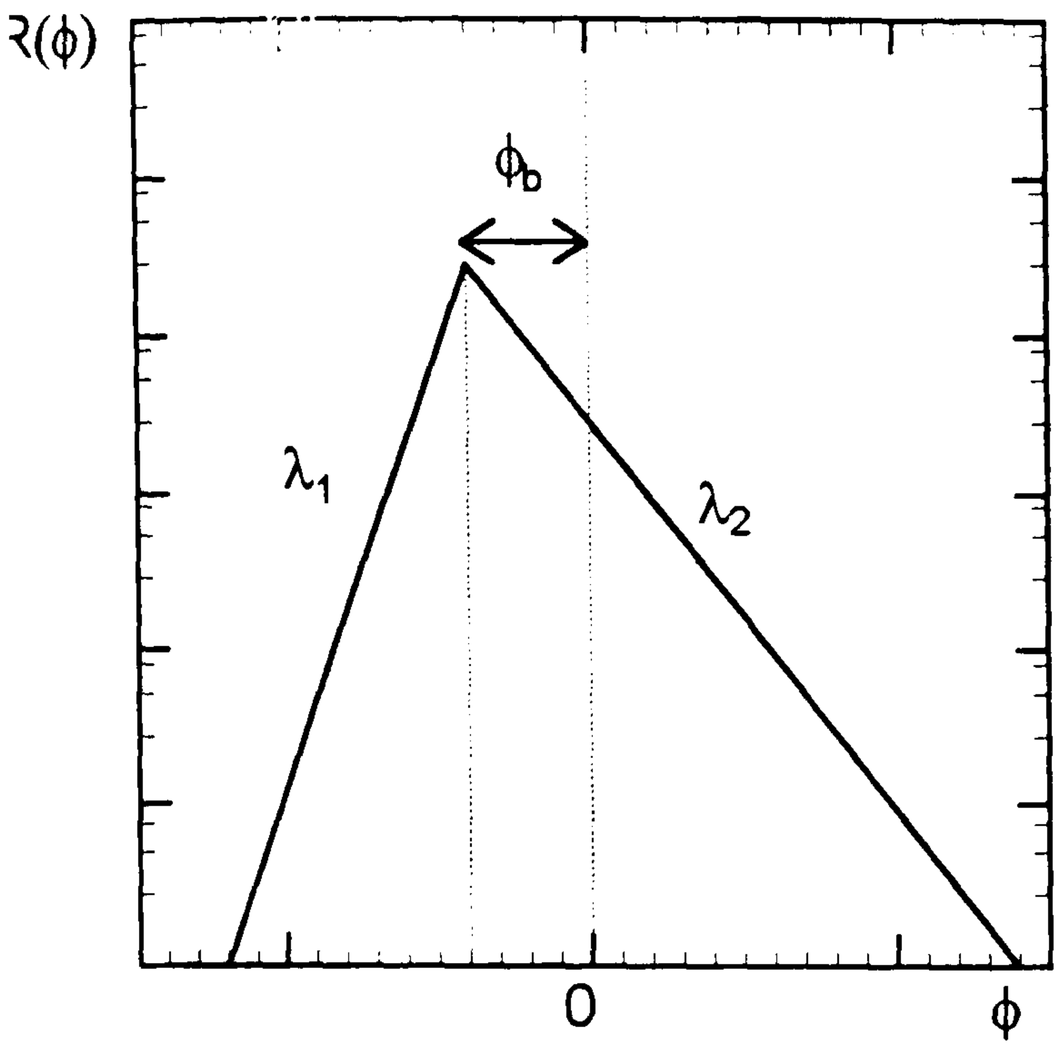
or, as the distribution is stored as set of discrete ordinates

$$H(x,W) = \sum_{j=1}^{N} (W_j - W_{j-1}) \log\left(\frac{x_j - x_{j-1}}{W_j - W_{j-1}}\right). \qquad (5.12)$$

5.1.2. An Analytic Response function.

To allow faster calculation of the convolution integral, an analytic response function was used. This approximate response function $R(\varphi)$ was

a double exponential, characterised by two slopes, λ_1 and λ_2, and a shift φ_b, which could be selected by the user.



$$R(\varphi) = \frac{1}{\lambda_1 + \lambda_2} e^{\varphi/\lambda_1}, \quad \varphi < \varphi_b, \tag{5.13}$$

$$R(\varphi) = \frac{1}{\lambda_1 + \lambda_2} e^{-\varphi/\lambda_2}, \quad \varphi \geq \varphi_b \tag{5.14}$$

Figure 5.3. The analytic response function.

5.1.3. The Objective function.

A misfit statistic must be used to compare the profile expected from the current trial distribution *x(W)* with the measured profile $Y(\varphi)$. The misfit statistic suggested by Skilling (section 3.2) is chi-squared:

$$\chi^2 = \sum_{\varphi} \frac{\left(Y_{calc}(\varphi) - Y(\varphi)\right)^2}{\sigma(\varphi)^2} \tag{5.15}$$

which is (catastrophically) undefined in the noiseless case $\sigma=0$. A statistic which does not have this shortcoming is the root mean square (r.m.s.) error e^2, given by

$$e^2 = \frac{1}{\varphi_{max}} \sum_{\varphi=0}^{\varphi_{max}} \left[Y_{calc}(\varphi) - Y(\varphi)\right]^2 \tag{5.16}$$

This form of misfit does not allow for the presence of experimental noise on the measured data, thus Collins included a term to account for additive background noise,

$$e^2 = \frac{1}{\varphi_{max}} \sum_{\varphi=0}^{\varphi_{max}} \left[\left(Y_{calc}(\varphi) + \sigma_b(\varphi) \right) - Y(\varphi) \right]^2, \qquad (5.17)$$

where $\sigma_b(\varphi)$ is the expected value of added background noise on each measured data point. As the concentration distribution and response are both normalised, the misfit is calculated on normalised SIMS (measured and simulated) data.

To attempt to simultaneously maximise the entropy *H* and reduce the misfit e^2, the standard method of Lagrangian (unknown) multipliers was to be used (section 3.3). The potential function to be used was a variant of that given by Skilling,

$$V = \left(\sigma^2 + \alpha \right) H + (1 - \alpha) e^2 \qquad (5.18)$$

where *H* is the Shannon entropy, σ the expected variance of the noise on the profile, and e^2 is the r.m.s. error. The variable α is the unknown multiplier, in this case it is to be chosen by the user.

5.1.4. The Search.

5.1.4.1. The Initial position.

In any optimisation problem, great increases in speed may be made by starting closer to the solution. However, choosing such a starting position reliably is not simple, thus in this algorithm the initial position chosen was equivalent to a uniform surface - the maximum entropy solution in the absence of data. As the algorithm was to begin with three blocks, the four ordinates were initially placed at $(x, w) = (0,0)$, $(\frac{1}{3}, \frac{1}{3})$, $(\frac{2}{3}, \frac{2}{3})$, and $(1,1)$.

5.1.4.2. Movement of the ordinates.

At the heart of the search algorithm was the concept that modern personal computers are quick enough that a simple search algorithm would suffice, at least in the initial development stages. Accordingly no use was made of a gradient vector or the Hessian (second derivative) matrix - each ordinate was simply tried in a number of positions. The extreme points (ordinates *0* and *N*) are fixed by the form of the polygonal arc at (0,0) and (1,1) respectively. For the other ordinates, the monotonic nature of the polygonal arc gives a region within which any motion may occur - the *bounding rectangle*. This rectangle was used to select the trial positions.

For each of the inner ordinates, ordinates *2* to *N-2*, two new positions were attempted, keeping the best of these two and the previous position. The new positions chosen were on the normal to the line between the nearest neighbours, a fraction τ of the distance between the current position and the bounding rectangle. In this way, the graph was moving approximately normal to the local tangent. Initially, τ was given the value 1/4; it was intended to find an optimum value during the development of the method.

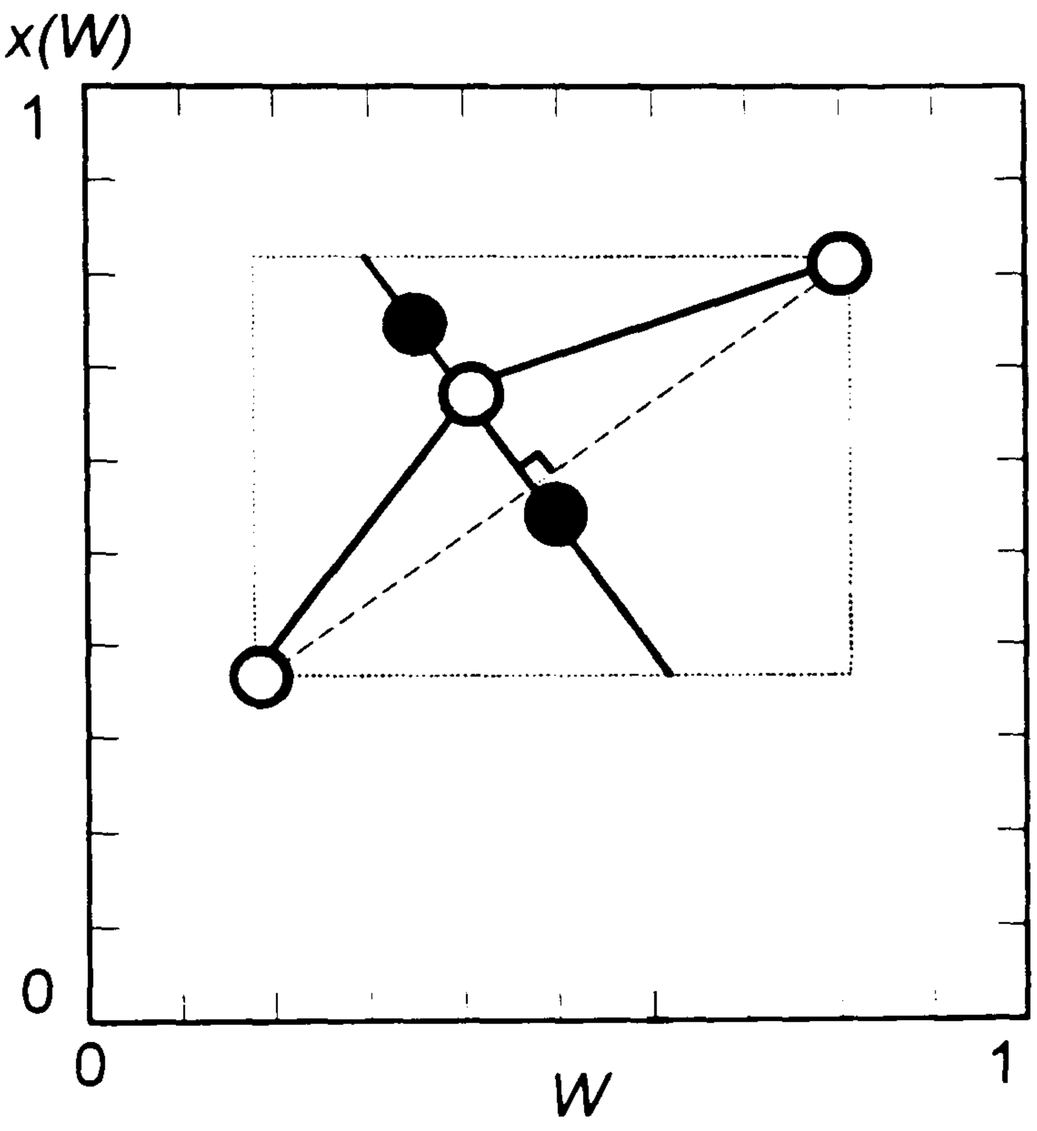


Figure 5.4. The bounding rectangle and two trial positions for an internal ordinate.

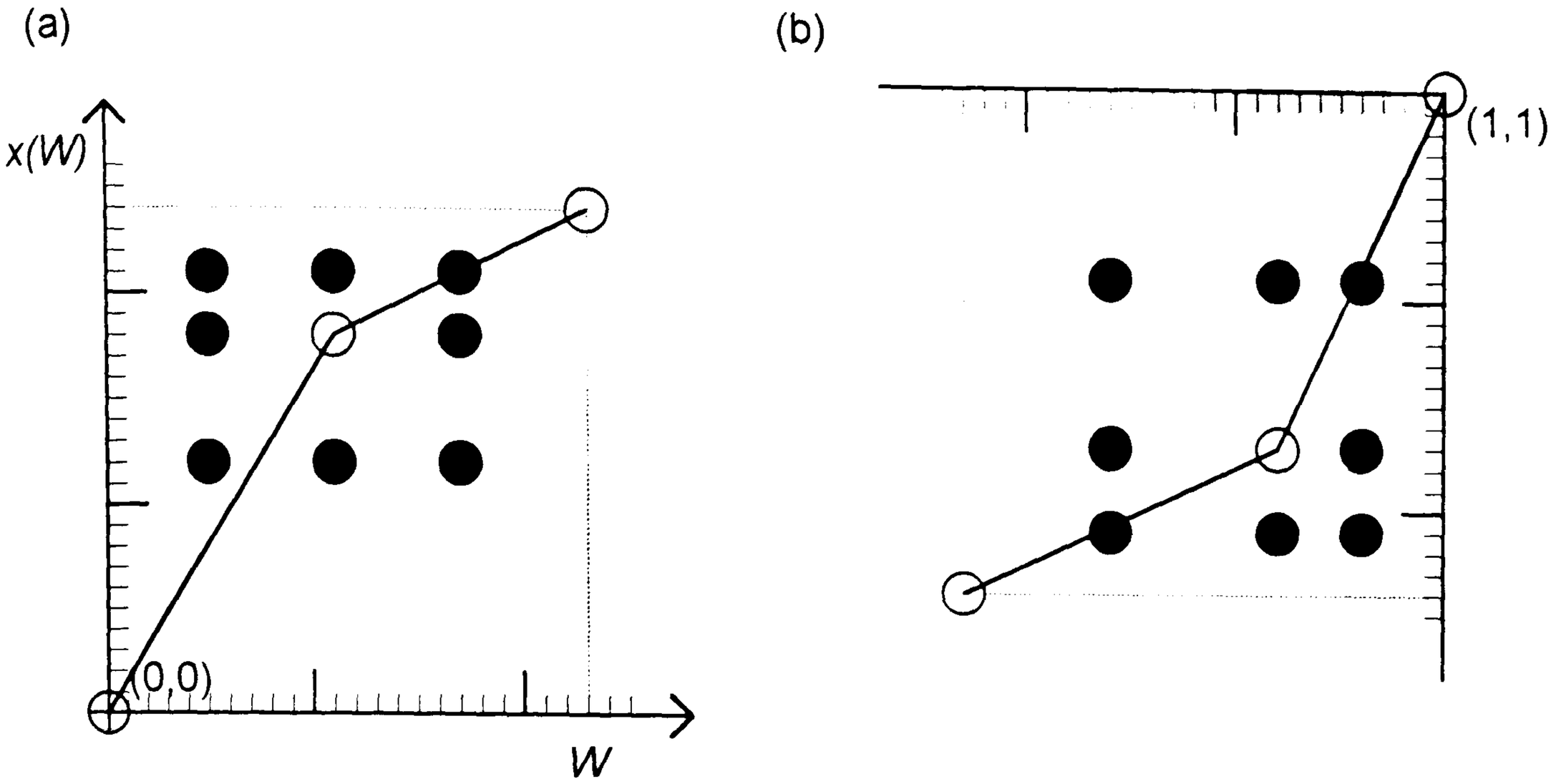


Figure 5.5. The bounding rectangles and eight trial positions for (a) the first, and (b) the last mobile ordinates.

Initial trials with the algorithm (Collins, private communication) had demonstrated that a different search procedure for the mobile outer ordinates (points *1* and *N-1*) reduced the time taken to converge. For these points *eight* new points were attempted, and the best of all nine points accepted as the new position. These points were all combinations of halfway to the bounding rectangle along or at 45° to the coordinate axes, as shown in figure 5.5.

5.1.4.3. Repeated motions: a pass.

The movement of all *N-2* mobile ordinates was termed a *pass*. At the end of every pass, progress was assessed: If an ordinate and its nearest neighbours had not moved successfully, any later attempts on this ordinate would necessarily have tried the same position. To avoid this repetition the ordinate was fixed. Passes were repeated until all the ordinates were fixed in this way.

5.1.4.4. Increasing the number of blocks.

At the end of a series of passes, once all the ordinates were stationary, the program was allowed to increase the number of blocks. Collins had devised a number of algorithms for doing so, the relative merits of each unknown:

(i) Every block in the histogram was divided in two. This has the effect of doubling the number of blocks and thus reaching the desired maximum number quite quickly.

(ii) The block of largest area was divided into three. This has the effect of concentrating the program on regions which may be under-resolved.

(iii) The two blocks on either side of the largest increase in W are spilt into two, concentrating the profile on the slopes of the peaks.

The method used in any reconstruction was to be chosen by the user, until such a time as repeated tests demonstrated that one method (or any other proposed method) was best.

5.1.4.5. Convergence criterion.

The algorithm had two termination criteria:

(i) If, at the end of a pass, no ordinate had been successfully moved, and the number of blocks was equal to the (user-specified) limit, successful convergence is assumed.

(ii) To give an upper limit to the time taken by the algorithm, if the number of trial position exceeded a limit, or the number of passes exceeded a pre-defined limit, execution was terminated, unconverged.

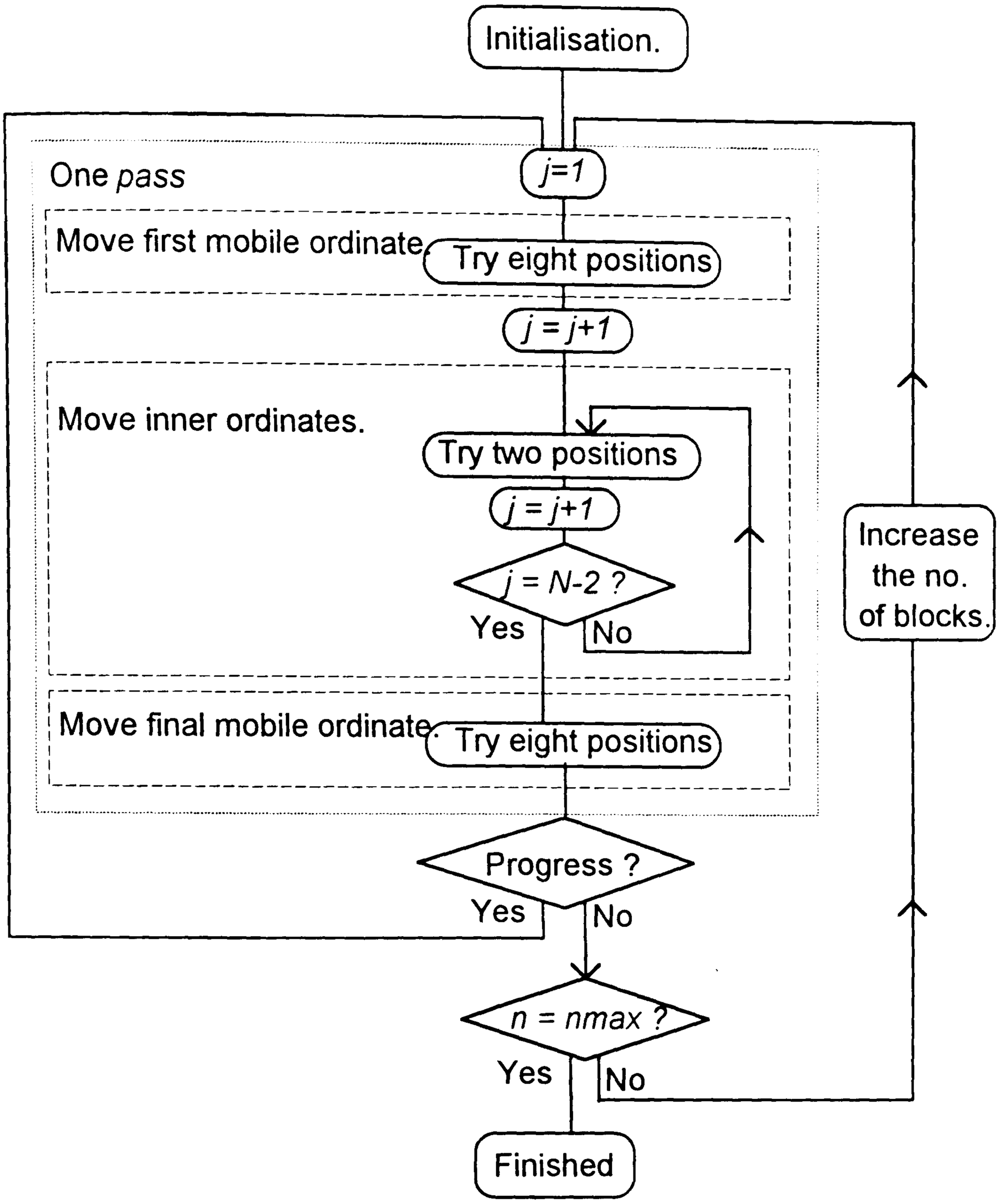


Figure 5.6. Flow chart for the histogram search of Collins.

5.1.5. Analysis of the algorithm.

Initially, the rate of convergence was too slow to perform the optimisation with more than some twenty blocks. To further develop Collins' approach, we decided it would be useful to determine effects of the search procedure *in terms of the changes to the concentration distribution C(z).*

5.1.5.1. The meaning of the bounding rectangle.

Indexing the ordinate to be moved k, and denoting the new position $\left(x_k^{'}, W_k^{'}\right)$, the boundary conditions may be written

$$x_{k-1} < x_k^{'} < x_{k+1} \qquad (5.19a)$$

$$W_{k-1} < W_k^{'} < W_{k+1} \qquad (5.19b)$$

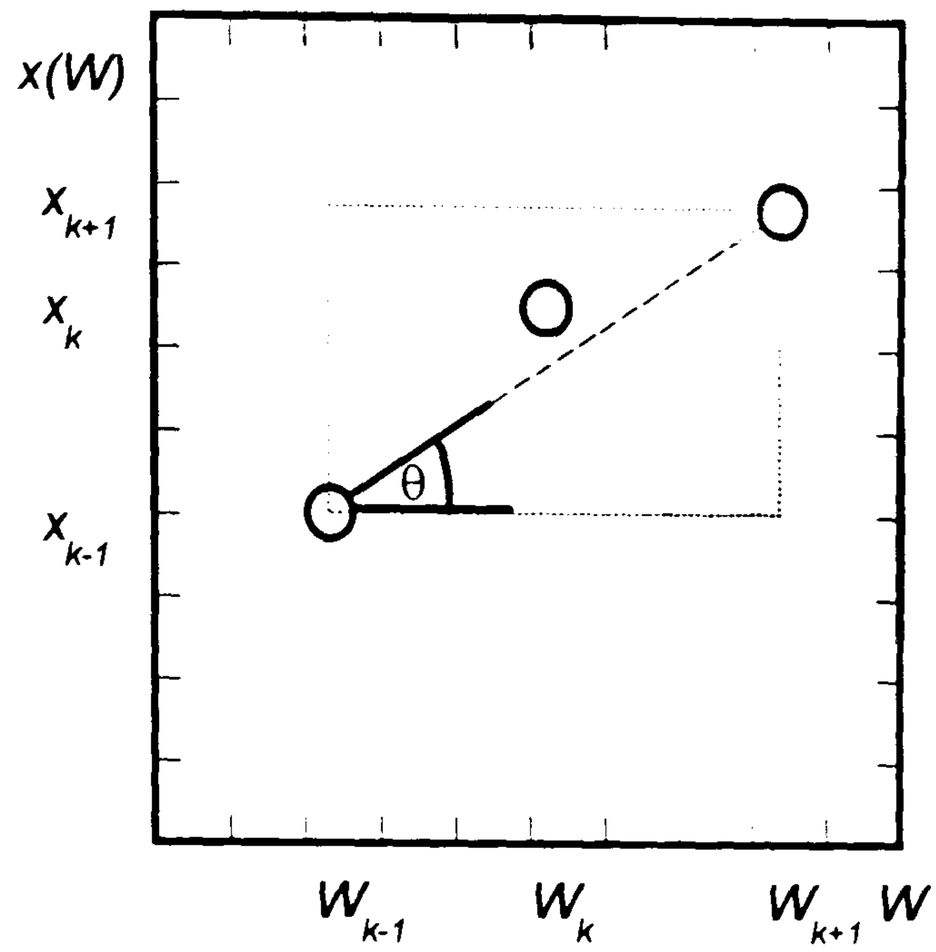


Figure 5.7. The bounding rectangle.

By subtracting W_{k-1}, which is invariant during the motion of point k, equation 5.19b may be rewritten

$$0 \quad < \quad w_k^{'}\left(x_k^{'} - x_{k-1}\right) \quad < \quad w_k^{'}\left(x_k^{'} - x_{k-1}\right) + w_{k+1}\left(x_{k+1} - x_k^{'}\right) \qquad (5.20)$$

Separating the boundary conditions into a depth-boundary (in terms of x), and a concentration boundary (in terms of w), it can be seen that the two boundary conditions are not active simultaneously. Defining the angle θ as in figure 5.7,

$$\tan\theta \quad = \frac{x_{k+1} - x_{k-1}}{W_{k+1} - W_{k-1}} \qquad (5.21)$$

$$= \frac{x_{k+1} - x_{k-1}}{w_k\left(x_k - x_{k-1}\right) + w_{k+1}\left(x_{k+1} - x_k\right)} \qquad (5.22)$$

Thus if $\theta<45^0$, the depth boundaries are active, and the motion of the polygonal arc is limited such that the *width* of both this block and its neighbours remains non-zero. If $\theta \geq 45^o$, the concentration boundaries are active, and the motion of the block is limited by the *areas* of this block and the

next: the areas of both blocks must remain non-zero. (The normalisation inherent in the approach will ensure that the total area of these two blocks is constant).

5.1.5.2. The motion of the internal points.

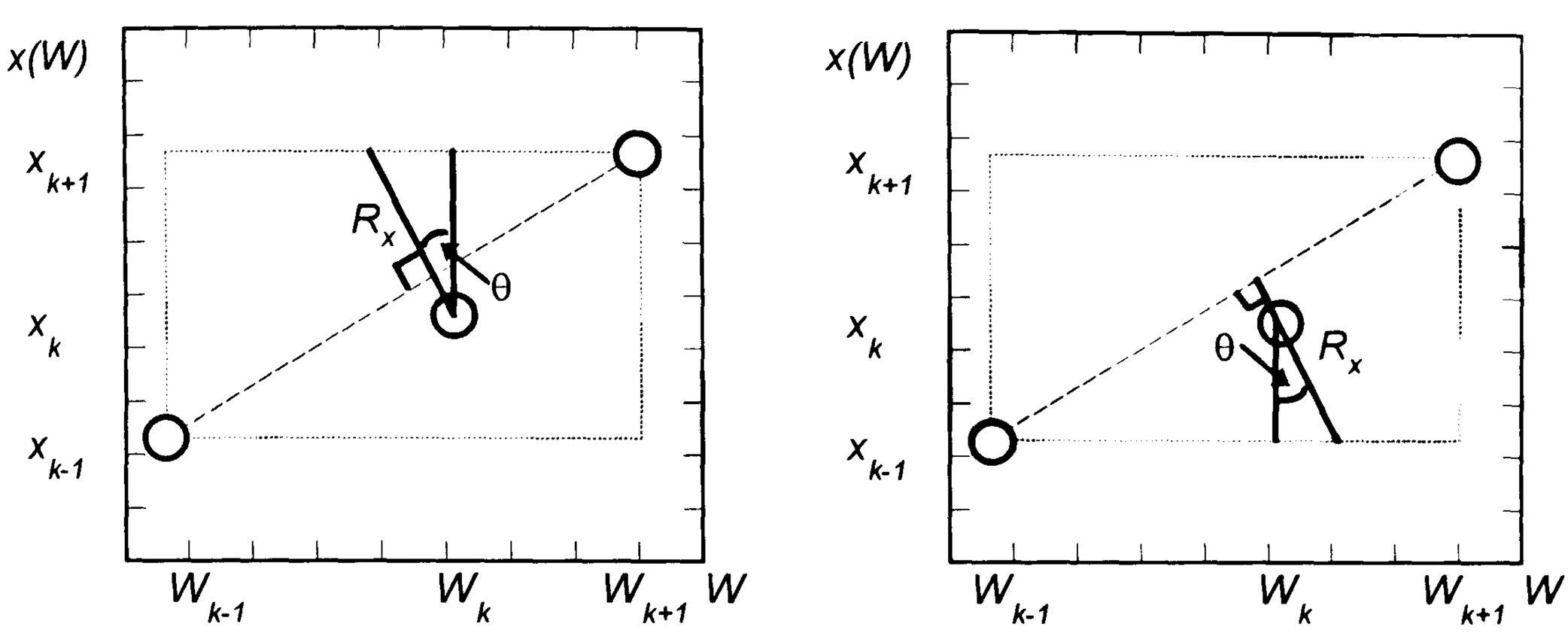


Figure 5.8. Motion of the polygonal arc (a) 'upward', and (b), 'downward'.

Looking now at the motion of the internal ordinates, we again find two types of behaviour, depending on the angle θ.

5.1.5.2.1. The case of $\theta<45^c$.

For angles less than 45°, the depth boundary limits the motion. Considering first the 'upward' motion of ordinate *k*, as shown in figure 5.8a, geometric considerations tell us that the distance to the *x*-boundary, R_x is

$$R_x = \frac{x_{k+1} - x_k}{\cos\theta} \tag{5.23}$$

Given that the point might be moved a fraction τ along this line, we deduce the new position $(x_k^{'}, W_k^{'})$ as

$$x_k^{'} = x_k + \tau(x_{k+1} - x_k), \tag{5.24a}$$

$$W_k^{'} = W_k - \tau(x_{k+1} - x_k)\tan\theta \tag{5.24b}$$

To obtain the new position in terms of the block height w, we subtract W_{k-1} from both sides of equation 5.24b, and rearrange to give

$$w_k'(x_k' - x_{k-1}) = w_k(x_k - x_{k-1}) - \tau(x_{k+1} - x_k)\tan\theta \qquad (5.24c)$$

thus

$$w_k' = \frac{w_k(x_k - x_{k-1}) - \tau(x_{k+1} - x_k)\tan\theta}{(x_x - x_{k-1}) - \tau(x_{k+1} - x_k)} \qquad (5.24d)$$

For the 'downward' motion, as shown in figure 5.8b, the range of motion to the lower x boundary is given by

$$R_x = \frac{x_k - x_{k-1}}{\cos\theta} \qquad (5.25)$$

and the corresponding new position is given by

$$x_k' = x_k - \tau(x_k - x_{k-1}) \qquad (5.26a)$$

$$W_k' = W_k + \tau(x_k - x_{k-1})\tan\theta, \qquad (5.26b)$$

thus

$$w_k' = \frac{w_k + \tau\tan\theta}{1 - \tau} \qquad (5.26c)$$

5.1.5.2.2. The case of $\theta \geq 45^0$.

If $\theta \geq 45^0$, the W boundary restricts the motion. Similar geometric considerations now lead to slightly different formulas. For 'upward' motion,

$$x_k' = x_k + \tau\left(\frac{W_k - W_{k-1}}{\tan\theta}\right) \qquad (5.27a)$$

$$= x_k + \tau\left(\frac{w_k(x_k - x_{k-1})}{\tan\theta}\right), \qquad (5.27b)$$

and

$$W_k' = W_k - \tau(W_k - W_{k-1}), \qquad (5.27c)$$

thus

$$w_k' = \frac{w_k(x_k - x_{k-1})(1-\tau)\tan\theta}{x_k \tan\theta + \tau w_k(x_k - x_{k-1})} \tag{5.27d}$$

For 'downward' motion,

$$x_k' = x_k - \tau\left(\frac{w_k(x_k - x_{k-1})}{\tan\theta}\right), \tag{5.28a}$$

$$W_k' = W_k + \tau(W_k - W_{k-1}), \tag{5.28b}$$

thus

$$w_k' = \frac{w_k(x_k - x_{k-1})(1+\tau)\tan\theta}{x_k \tan\theta - \tau w_k(x_k - x_{k-1})} \tag{5.28c}$$

Using these equations we may interpret the up/down motion of the polygonal arc as follows: an 'upward' motion of the arc at point *k* corresponds to making block *k* broader and lower, and the area of the block is decreased by the motion. Due to the normalisation, block *k+1* must compensate for this motion: although both W_{k+1} and x_{k+1} are constant, block *k+1* becomes taller and thinner, its area increasing such that the total area of the two blocks is constant. The size of the motion depends on either the width of the next block (equation 5.24) or the area of this block (equation 5.27), whichever is the minimum. This motion is shown schematically in figure 5.9a.

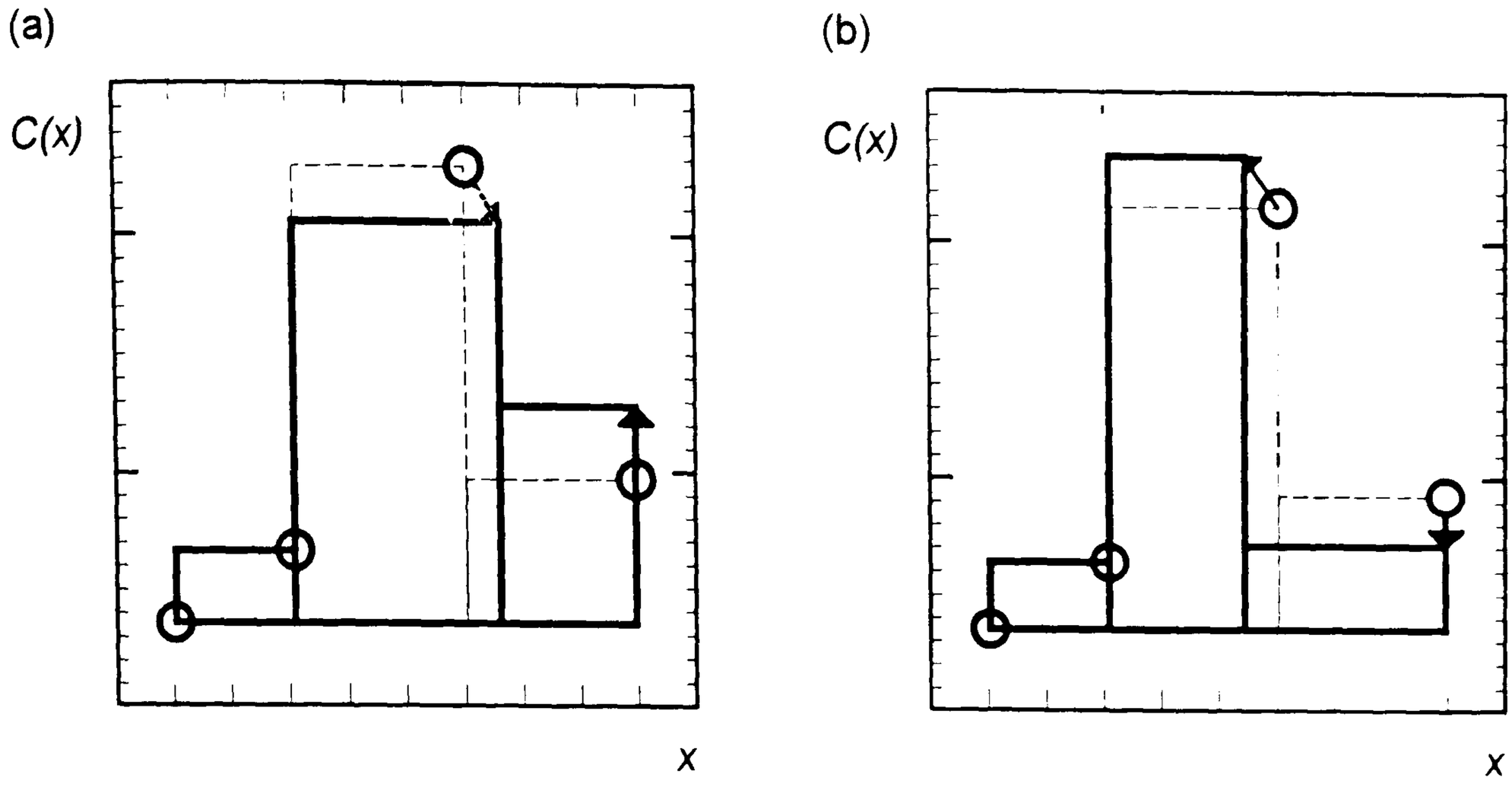


Figure 5.9. The equivalent positions for (a) 'upward' and (b) 'downward' motion of the polygonal arc.

The 'downward' motion of the arc corresponds to making block *k* taller and thinner, with larger area. The compensating motion of block *k+1* is to become lower and broader with smaller area, as shown in figure 5.9b. Here, the size of the step is proportional to the width of this block (equation 5.26) or the area of the next (equation 5.28).

It is now obvious why a different search procedure was adopted for the extreme points. With regard to figure 5.9, it can be seen that there is no way that a block may become both thinner and shorter simultaneously: it must do so by alternating between the diagonal motions when it is moved, and the compensating vertical motions when the previous block is moved. For the first block, there is no previous block to move, thus no such motion can occur.

5.1.5.3. The motion of the extreme points.

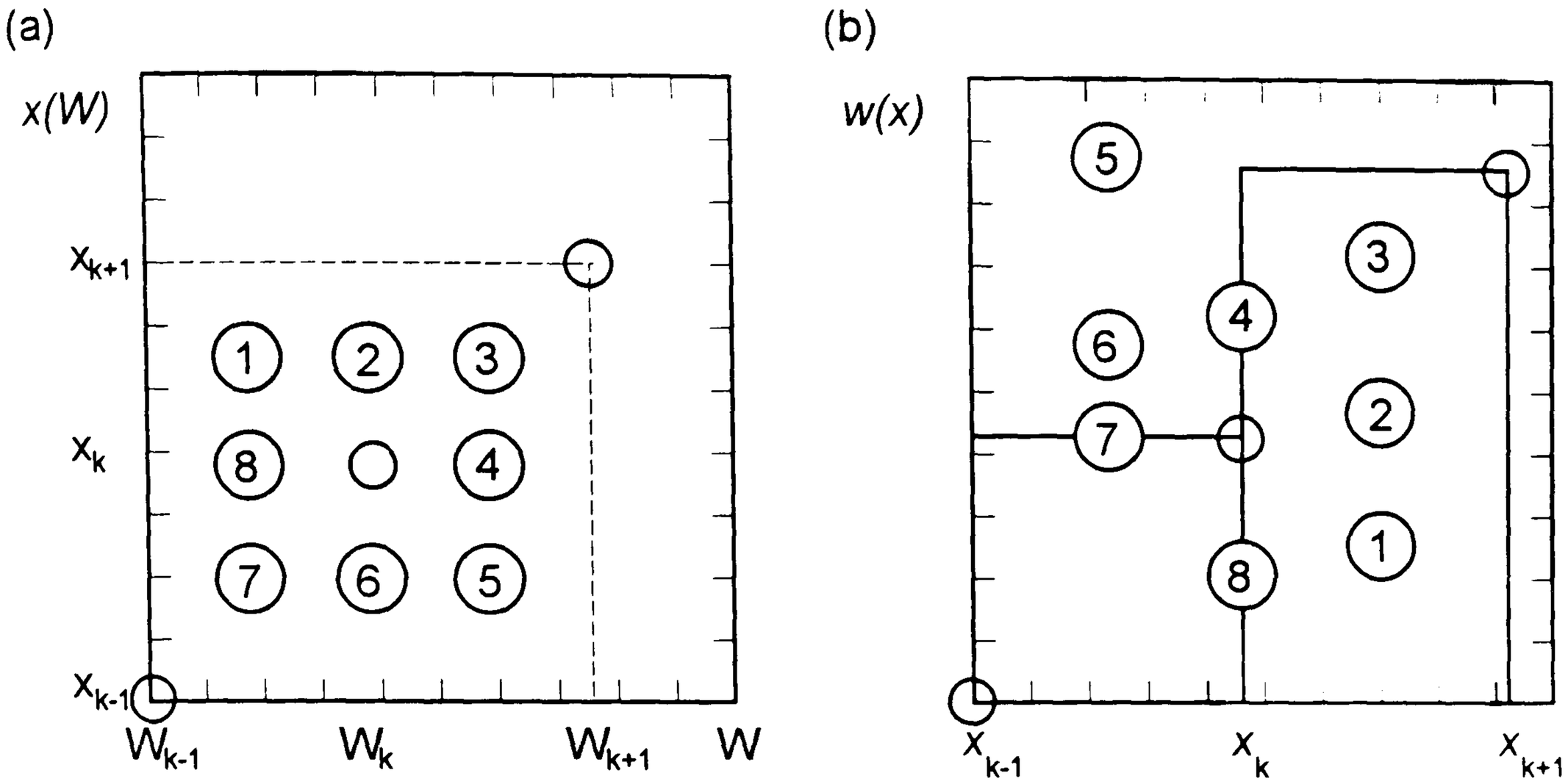


Figure 5.10. (a) The labelling of the eight trial positions for an extreme ordinate, and (b) the equivalent trial positions for the motion.

If the eight trial positions for the ordinate are labelled as shown in figure 5.10a, we may again interpret these motions in terms of the concentration distribution *w(x)*. The new positions are as shown in Table 5.1, in terms of the area A_j and thickness T_j of block *j*, where this clarifies the notation, and schematically in figure 5.10b.

Ordinate	New x_k	New w_k
1	$\frac{1}{2}(x_k + x_{k+1})$	$\frac{A_k}{2T_k + T_{k+1}}$
2	$\frac{1}{2}(x_k + x_{k+1})$	$2.\frac{A_k}{2T_k + T_{k+1}}$
3	$\frac{1}{2}(x_k + x_{k+1})$	$\frac{2A_k + A_{k+1}}{2T_k + T_{k+1}}$
4	x_k	$w_k + \frac{A_{k+1}}{2T_k}$
5	$\frac{1}{2}(x_{k-1} + x_k)$	$2\left(w_k + \frac{A_{k+1}}{2T_k}\right)$
6	$\frac{1}{2}(x_{k-1} + x_k)$	$2w_k$
7	$\frac{1}{2}(x_{k-1} + x_k)$	w_k
8	x_k	$\frac{w_k}{2}$

Table 5.1. The eight new positions for a mobile extreme point.

5.2. The Univariate Search.

The analysis of the histogram method had indicated that some of the concepts underlying the method gave little real advantage. The transformation to a monotonic function and the use of the bounding rectangle were also considered of questionable value. Attempts by the author to develop the algorithm quickly gave rise to very large numbers of variant algorithms; the assessment of each would have been extremely time-consuming. It was decided that a fundamental change to a more efficient

algorithm could potentially give results much more quickly than fine-tuning the histogram method. Accordingly a second algorithm was developed.

5.2.1. Basic concepts.

For this second algorithm, the rate of convergence was not deemed as important as the *rate of development of the program*. The basic design of this second algorithm was repeatedly simplified to reduce the number of variants that the algorithm could spawn during development. Other factors considered important during the design stages were:

(i) **No Normalisation.** The normalisation of the concentration distribution must be removed: the results of the optimisation should be quantitative with as little further processing as possible.

(ii) **No implicit assumptions about the depth mapping.** The algorithm should perform the reconstruction entirely as a functions of frame or primary ion dose, leaving the depth mapping separate, as in section 4.1.

(iii) **Speed of the convolution is paramount.** As the majority of the time is spent calculating the convolution integral, any proposed transformations of the variables should speed the convolution.

(iv) **The removal of time limits.** There should be no limitations on the time allowed for an optimisation: such a process merely yields unconverged results.

5.2.2. Representation of the distribution.

In this algorithm the concentration distribution is stored as a list of concentration values. This form was chosen for a number of reasons. Firstly, as the spacing of the data points is fixed (preferably at a convenient unit such

as the frame number) each ordinate has effectively only one variable, resulting in faster convergence than the histogram method for an equivalent resolution. Secondly, the convolution equation may be evaluated using the property of the Fourier transform (section 3.1) The computational cost of these transforms is greatly reduced by use of the Fast Fourier Transform (FFT) (Brigham, 1988). Finally, this equal-spaced representation is identical to that of the normal SIMS data, thus the results of the optimisation will not have the unfamiliar look of a general histogram.

5.2.3.The objective function.

5.2.3.1. The calculation of the entropy.

The general formula for the Shannon-Jaynes entropy, for discrete distributions,

$$H(C|M) = \sum_{j=1}^{N} C(j) - M(j) - C(j)\log\left(\frac{C(j)}{M(j)}\right) \qquad (5.29)$$

can be greatly simplified if both the distribution *C(f)* and the model *M(f)* are normalised. Assuming that the model M*(f)* does not change during the optimisation it may be normalised at the start of the program with little computational cost. However, *C(f)* is to be represented *without* normalisation, thus the normalisation must be performed *as the entropy is calculated.* Accordingly, defining

$$K = \sum_{j=1}^{N} C(j) \qquad (5.30)$$

$$H(C,M) = -\sum_{j=1}^{N}\left[\frac{C(j)}{K}\log\left(\frac{C(j)}{KM(j)}\right)\right] \qquad (5.31)$$

Initially, as before, we shall let the model be uniform,

$$M(j) = 1 \qquad j = 1..N \tag{5.32}$$

to allow the data to speak without prejudice.

This form of the entropy will be costly to compute, as the computation of the factor K will require an extra loop in the calculation of the entropy. To speed the computation we note that

$$H(C) = -\sum_{j=1}^{N}\left[\frac{C(j)}{K}\log\left(\frac{C(j)}{K}\right)\right] \tag{5.33}$$

$$= -\left[\sum_{j=1}^{N}\frac{C(j)}{K}\log(C(j)) - \sum_{j=1}^{N}\frac{C(j)}{K}\log(K)\right] \tag{5.34}$$

$$= -\left[\frac{1}{K}\sum_{j=1}^{N}C(j)\log(C(j)) - \frac{\log(K)}{K}\sum_{j=1}^{N}C(j)\right] \tag{5.35}$$

$$= -\frac{1}{K}\sum_{j=1}^{N}C(j)\log(C(j)) + \log(K) \tag{5.36}$$

thus the computation may be performed using a single loop. Any further adjustment to speed the calculation of the entropy would be superfluous, as the evaluation of the convolution integral at each stage, even using the FFT, is far slower.

5.2.3.2. The misfit statistic.

With the convolution model adjusted to account for any background signal (section 4.4), the remaining difference between the calculated and measured profiles will be, at least approximately, of Gaussian form (section 4.4). Accordingly for this algorithm we use the misfit statistic suggested by Skilling, equation 3.20.

$$\chi^2 = \sum_{j=0}^{N}\frac{\left(Y_{calc}(j) - Y(j)\right)^2}{\sigma(j)^2} \tag{5.37}$$

The potential catastrophic errors in the noiseless case may be avoided by a simple 'if' statement during the calculation of σ, replacing zero noise with an arbitrary small value. For genuine data noise will always be present.

5.2.3.3. The Lagrangian multiplier.

As before the misfit was combined with the Shannon-Jaynes entropy using a single Lagrangian multiplier α, giving an objective function

$$V = \alpha H - \chi^2 \tag{5.38}$$

as with the more recent probabilistic derivations of the maximum entropy method. In contrast with the histogram method, however, the value of α was to be selected such that the final misfit was equal to the number of data points, equivalent to a single standard deviation mismatch in each frame (The historic MaxEnt method, section 3.3.1). Obviously, this requires repeated use of the optimisation search.

5.2.4. The search algorithm

With the value of α determined as for the historic MaxEnt method, an optimisation of the potential function V is required for each value of α. Accordingly the search must be very efficient. Nonetheless, true to the ideas of the initial optimisation method the code developed uses no gradient or Hessian calculations. The algorithm is a simple univariate search, moving a single ordinate at a time. It was found useful initially to develop the algorithm without the convolution, even though MaxEnt deals rather poorly with such un-blurred data (Gull, 1988), yielding always a single standard deviation toward the mean.

There are essentially four components to this search: the selection of an initial position; the motion of each ordinate; the method of refinement; and the convergence criterion. These will be described in order.

5.2.4.1. The initial position.

If the search is functioning correctly, the final solution is not affected by the initial position (or the route taken). The initial position may affect the time required to converge as it changes the distance to the maximum. Accordingly a number of possible initial positions were considered and assessed.

5.2.4.1.1. A uniform starting position.

The simplest initial position is, as with the histogram method, a uniform surface. However, in this non-normalised treatment the height of the surface is a variable. A sensible first approximation was made: the initial position is such that the integrated concentration is correct. Accordingly, we select

$$C_{init} = \frac{\sum Y}{\sum R} \tag{5.39}$$

This initial position was to be the basis by which all others were measured.

5.2.4.1.2. Using the shape of the measured depth profile.

The second method was to accepting the SIMS data as a first approximation to the *shape* of the distribution. We thus select

$$C_{init}(j) = \frac{Y(j)}{\sum R} \tag{5.40}$$

This choice was not found to yield a consistent speed improvement. It was concluded that this was due to the shifting nature of the response function: the algorithm may only move ordinates vertically, and a small horizontal shift may cause a large vertical change *in the wrong direction*. This is shown in figure 5.11, where the uniform surface (dotted line) is closer to the 'true' profile (solid line) than a shifted profile (dashed).

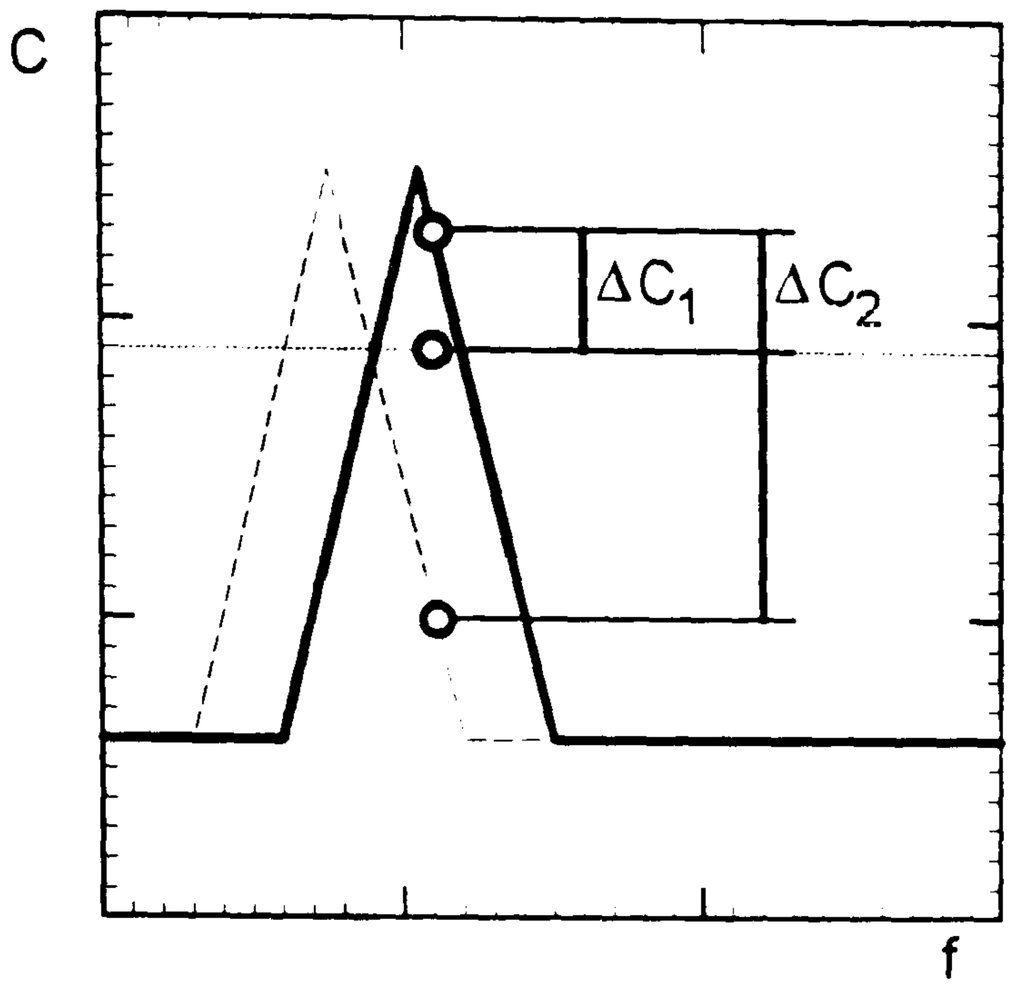


Figure 5.11. Distance of a profile from a uniform surface, and from a shifted profile.

5.2.4.1.3. Using an inverse Fourier deconvolution.

It was then considered that the solution of a Fourier deconvolution (section 3.1) might be closer to the MaxEnt solution than the measured SIMS data, as the deconvolution would perform some of the work of the reconstruction. However, the amplification of the noise leads to negative values. These are automatically further from the solution than a (positive) uniform surface. Two methods were attempted to make the Fourier-deconvolution initial position more effective:

(i) Replacing negative concentration values with a small, positive value. This was (not surprisingly) found to be particularly inefficient: the profile obtained is extremely irregular.

(ii) Smoothing the initial position. This was found to be the more effective, and did demonstrate a considerable speed improvement (up to 20%

fewer convolutions than the basic method) for some test samples. The optimum starting position appeared to be a 5-point running-average smoothed Fourier deconvolution. Still, in some test cases the algorithm was actually slowed by this method. As stated in chapter 3, the smoothed inverse Fourier deconvolution is not consistent with the measured data. It was concluded that in some cases it may even be 'further' from the MaxEnt solution than a uniform surface !

5.2.4.1.4. Retaining the optimum position for a previous value of α.

Finally, the starting position attempted was to use the results of the MaxEnt reconstruction from a previous value of α as the initial position for this value of α. This was considered a plausible choice because the MaxEnt solution does not suffer the logical inconsistency problem of the smoothed Fourier deconvolution. Obviously, this method is not possible for the first value of α. In this case the simplest of the above methods was selected (method 5.2.4.1.1). This initial position has been found to be very effective. The optimisation for the second and later values of α were found to require up to 25% fewer convolutions than the basic method.

A further increase in speed might be obtained using this method: with the initial position rather close to the optimum, the use of large step sizes is very wasteful ($2N$ convolutions are required to reduce the step size). However, determining the new optimum step size is non-trivial, thus the added complexity was not considered worthwhile.

5.2.4.2. The motion of the ordinates.

The representation of the distribution as a set of ordinates limits the possible search directions to two per point; it can go up or down (or remain at

its current location). As all variables have the same dimensions a single type of motion may be used for all. There is now no need to discriminate between 'interior' and 'exterior' points, and there is no need for stationary ordinates. The motion of each point was kept simple - a fixed step was to be taken in each direction (giving an inexact line search in each coordinate direction). As up and down are mutually exclusive it is not always necessary to try both; if up has yielded an increase in the potential function, down will not. A useful feature of many other optimisation algorithms was incorporated - once a direction had proved to be the correct one, steps in this direction were continued until the motion was exhausted. Acceleration along the search direction was not used, however, as initial tests showed no consistent improvement in speed.

Retaining the terminology that one complete set of motions is a pass, the convergence is examined at the end of each pass. Passes are repeated until no ordinate is successfully moved at this step size. An important difference between this algorithm and the histogram method is that a failure to move a point in one pass does not fix that point in future passes: the nature of the convolution allows interaction between different ordinates and to assume that no interaction exists is to court disaster.

5.2.4.3. Refining the step size.

The univariate search refines itself by reducing the size of the step taken. The step size ΔC is reduced (divided by a constant, ξ) whenever a pass results in no successful motions. Another set of passes is then started.

The initial step size, ΔC_{init}, was selected such that a single downward step would reach the minimum possible value of 1 atom cm^{-2} i.e. C_{init}-*1*.

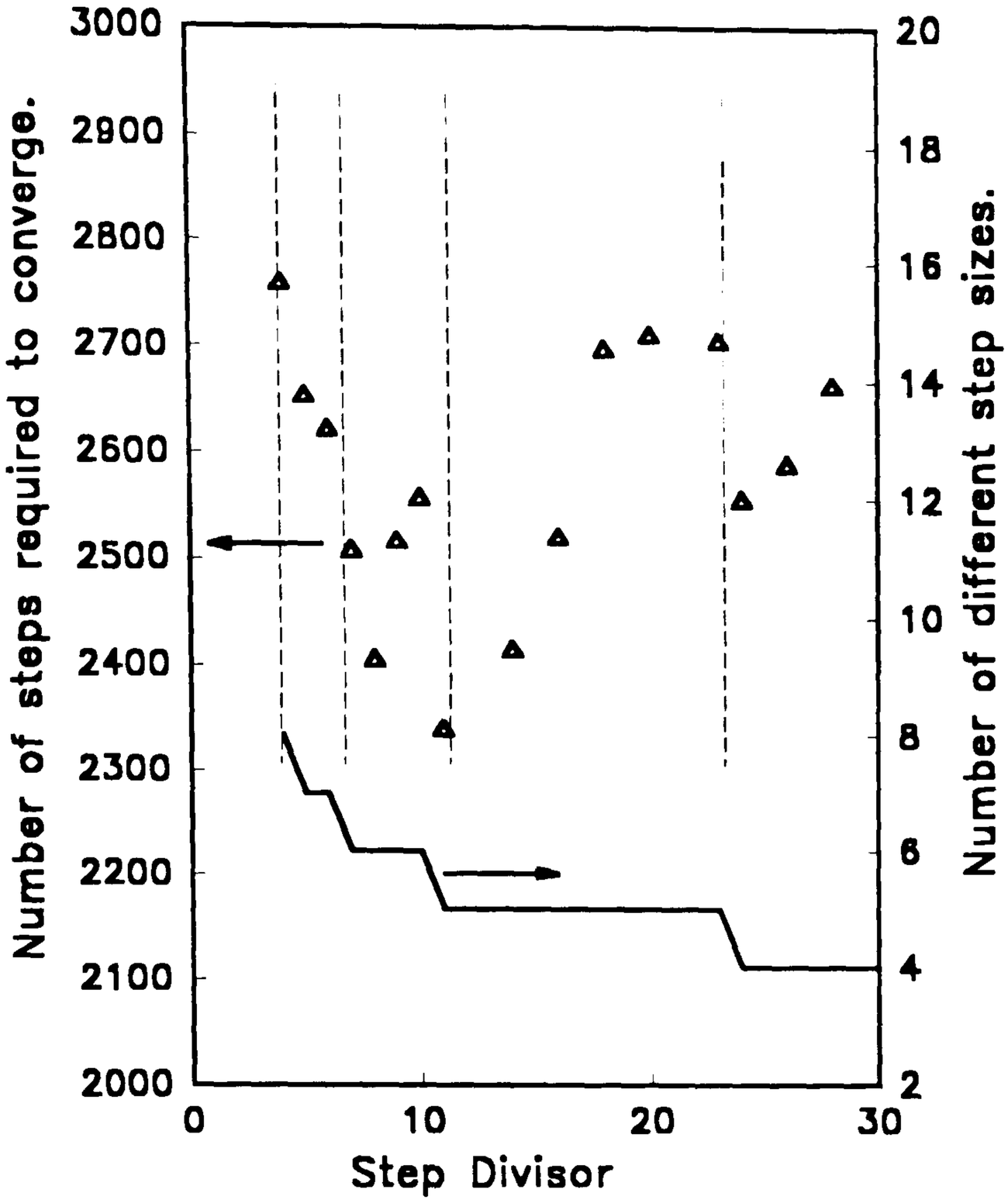


Figure 5.12. The effect of the step size divisor on the speed of the program.

For a correctly functioning search, the amount by which the step size is reduced does not affect the final solution, it merely affects the speed of convergence: If the step size is reduced only a little it will still not be possible to move any ordinates; reduced too much and many small steps will be required. A sensible range for this divisor was thought to be between 2 and 100, and the initial (arbitrary) choice was 10. An investigation of the optimum value was performed. The results for one set of test data are shown in figure 5.12. The behaviour may be explained as follows: The initial step size is determined by the data, the final step size specified by the user. For any chosen value of the divisor ξ, there will be a number of different step *sizes* used. The *n*th step size is given by

$$\Delta C_n = \frac{\Delta C_{init}}{\xi^{n-1}} \qquad (5.41)$$

The changes in efficiency depend on whether the step sizes are evenly spaced in the range. If, for a value of ξ, the nth step size is only slightly larger than ΔC_{final}, the algorithm must still attempt to use ΔC_{final}, as instructed. The final divisor thus approaches 1, and the final step size is wasted. If the divisor in increased, the amount of wasted steps increases, until the divisor is large enough that fewer step *sizes* are used, resulting in some saving. The divisor value may be efficiently selected by inverting equation 5.41,

$$\xi = e^{\left[\frac{1}{n-1}\log\left(\frac{\Delta C_{init}}{\Delta C_{final}}\right)\right]} \tag{5.42}$$

Alternatively, efficient use of the different step sizes can be ensured by rounding the initial step size up and the final step size down to the nearest power of ten, and always dividing by ten. As this method is simpler, it was adopted.

5.2.4.4. Convergence Criterion.

The simplicity of the algorithm yields an intuitive convergence criterion: The user specifies a minimum step size of interest (or of physical meaning). When a position is reached where no motions of this step size can increase the objective function, the program is deemed to have converged.

It should be remembered that univariate searches have a serious limitation if step sizes are too large: motion along non-coordinate directions can only be approximated using small step sizes. The use of such small step sizes may require many convolutions to converge, thus it is important that no limit is placed on the number of convolutions which can be performed. Provided a suitably small final step size is used, the program *cannot* terminate before the optimum is reached; it may, however, continue the optimisation indefinitely.

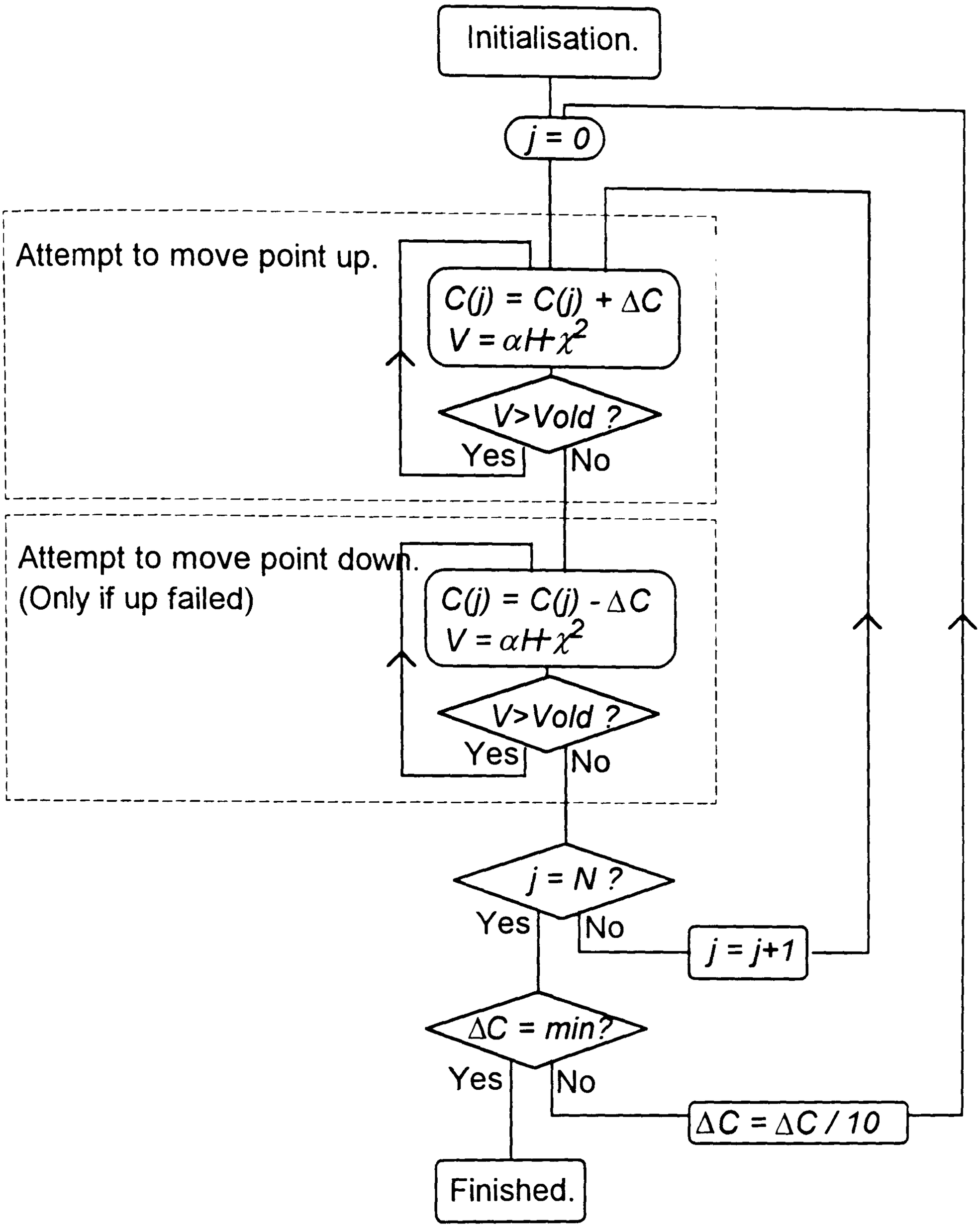


Figure 5.13. Flow chart for the univariate search.

5.2.5. The value of the Lagrangian multiplier.

For a fixed value of α, the optimisation of V will result in an 'optimum' distribution $C(j)$, with a corresponding value of entropy H and misfit χ^2. If the optimisation has converged, we expect both the misfit and the entropy value to depend monotonically on the value of α: as it is increased, the relative importance of the entropy is increased, thus the solution tends toward large values of entropy and a correspondingly large mismatch. As α is reduced, we expect solutions with less entropy but a better match to the data, as shown schematically in figure 5.14. Accordingly, we find α by trial-and-error as follows.

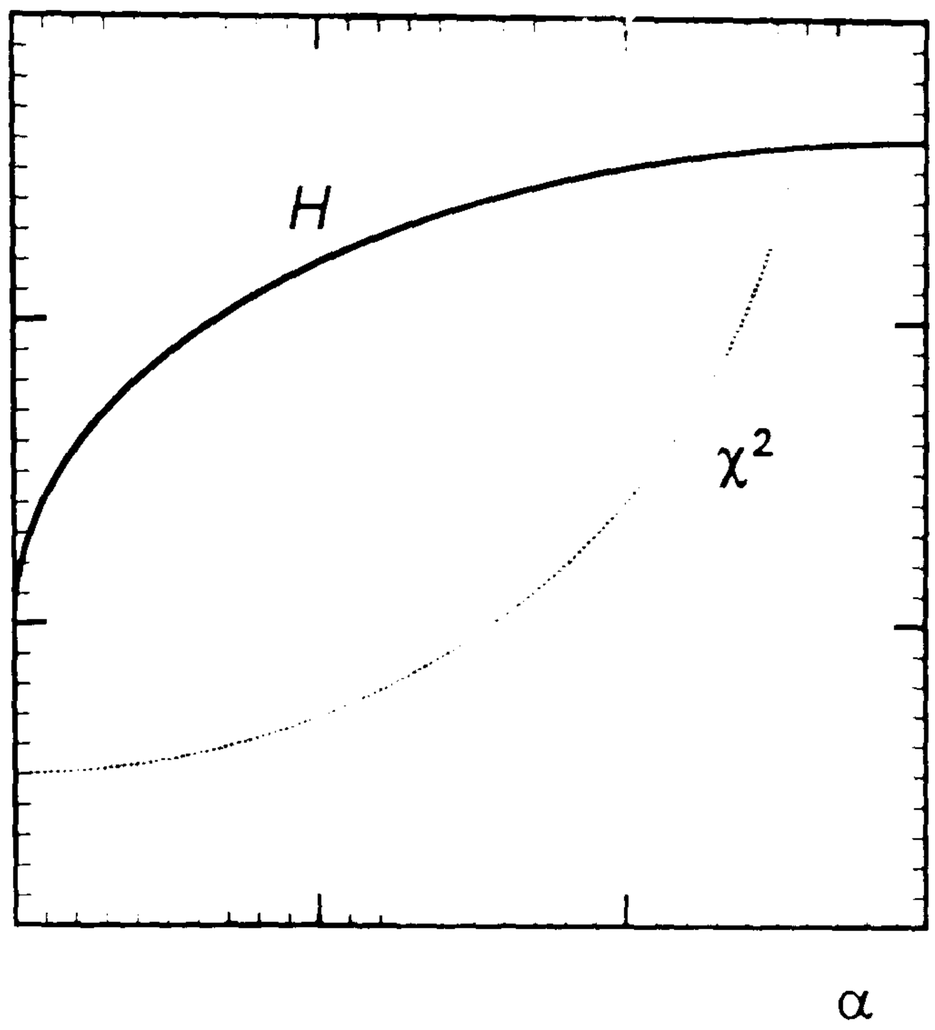


Figure 5.14. Schematic diagram of the dependence of H and χ^2 on α.

5.2.5.1. An initial check.

Initially, we perform the optimisation for the case $\alpha=0$. If the program is unable to find a solution such that $\chi^2 << N$, no other value of α will succeed: the SIMS data, response function, and noise estimate are inconsistent. This situation has occurred a number of times during this research - it is a very good indication of user error.

5.2.5.2. Finding a bounding α range.

If this first optimisation has succeeded, it is presumed that the data set is consistent and we attempt to find an optimum solution for non-zero α. The

first value of α selected will not matter greatly, and an arbitrary value of 500 was selected after initial trials. After each optimisation is performed, the value of χ^2 determines the next value of α. If $\chi^2<N$, we require a larger value of α, otherwise we require a smaller value. In this primary stage, the value of α is doubled until a value is found such that $\chi^2>N$.

5.2.5.3. Bisecting the range.

Given two α values such that the true value is somewhere between the two, the range is reduced by bisection; the optimisation is performed for an α halfway between the two limits and, depending on the value of χ^2, the upper or lower range is replaced with the middle value. This continues until the optimisation yields $\chi^2=N$ to within a user specified tolerance. It was decided that 0.1% was good enough, as the entropy was normally found to be a slowly varying function of α. The process is shown in figure 5.15.

5.2.5.4. Interpolation.

An alternative to the α-bisection method was developed. After the initial accelerating search has bounded the range of possible α values, the monotonically increasing nature of $\chi^2(\alpha)$ allows the use of an interpolation procedure. Three values of χ^2 corresponding to three values of α may be used to define a quadratic approximation to the function,

$$\chi^2(\alpha) \approx A\alpha^2 + B\alpha + C \tag{5.43}$$

where the coefficients are found by inverting the matrix equation

$$\begin{pmatrix} \chi_1^2 \\ \chi_2^2 \\ \chi_3^2 \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix} \begin{pmatrix} C \\ B \\ A \end{pmatrix} \tag{5.44}$$

Given the coefficients we may now find α such that

$$\chi^2 - N = 0 \tag{5.45}$$

using the standard formula

$$\alpha = \frac{-B \pm \sqrt{B^2 - 4A(C-N)}}{2A} \tag{5.46}$$

The greater of the roots will be the desired approximation to the maximum, the other will be outside the range (less than the minimum of the range).

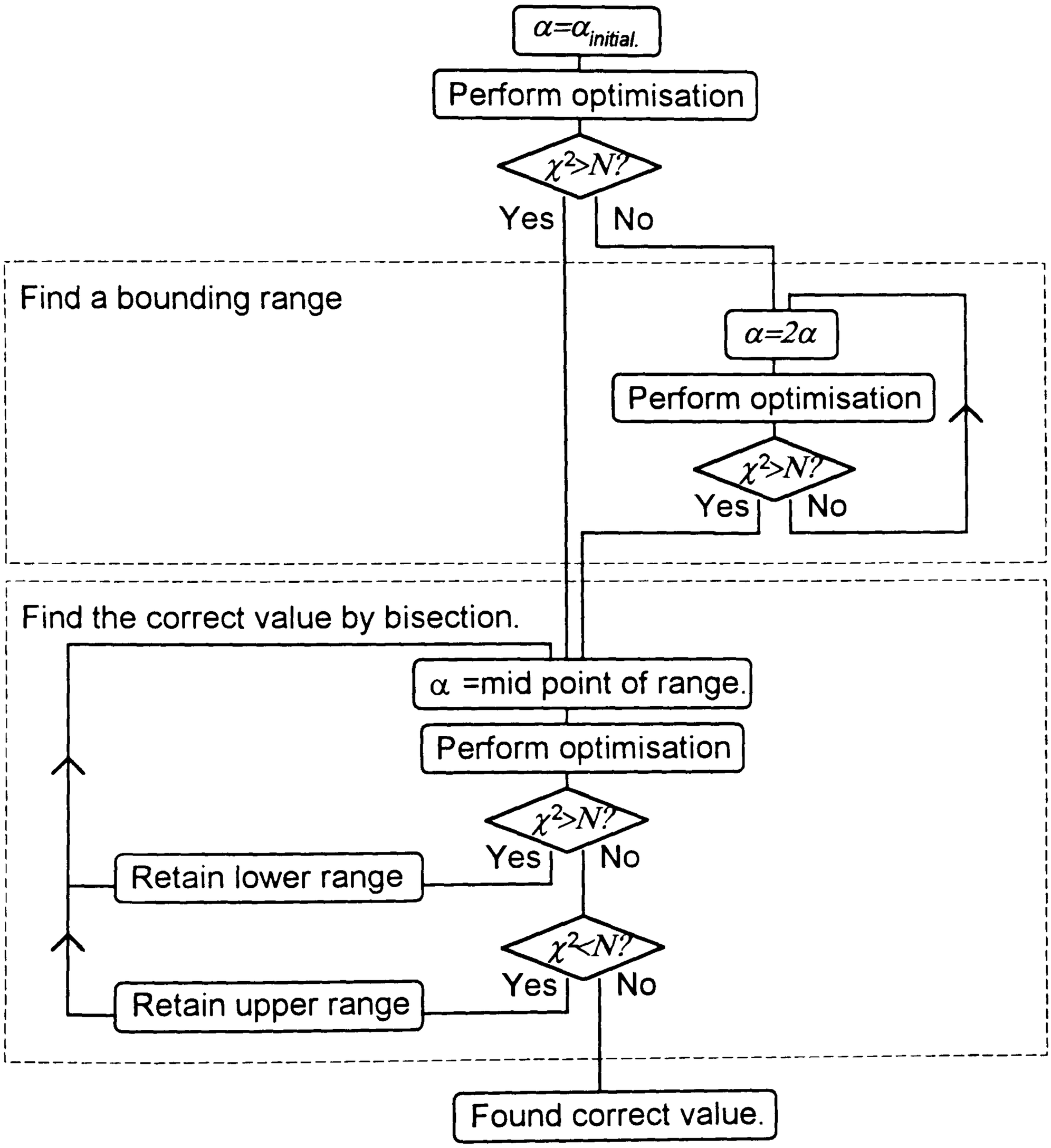


Figure 5.15. Flow chart for the α bisection algorithm.

5.3. The method of Davies, Swann and Campey.

Univariate searches have a number of disadvantages. Firstly, they can be forced to approach the maximum in an *N*-dimensional spiral pattern, rather than moving directly toward it. This inefficient behaviour is shown schematically in figure 5.16a for the two-dimensional case. Another type of inefficient behaviour is shown in figure 5.16b, where the univariate search must use small steps in alternate directions to emulate motion along a diagonal direction. These inefficiencies lead to slow convergence and thus may put a limit on the number of variables the algorithm can realistically optimise. More disturbingly, however, geometric features such as ridges may force a univariate search to terminate early. This would result in an unconverged solution.

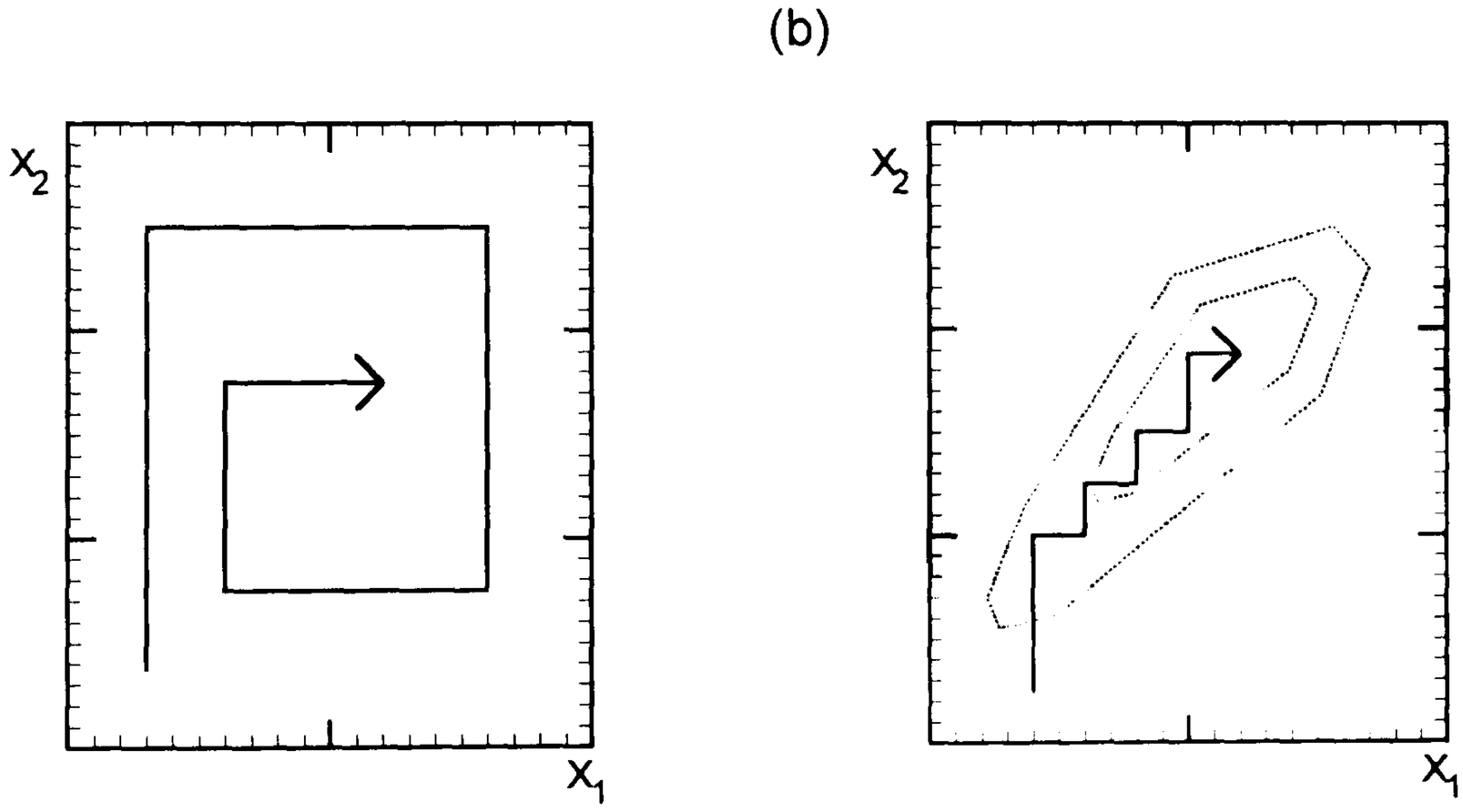


Figure 5.16. Possible inefficient routes take by a univariate search algorithm when the contours of the potential function are skew to the coordinate directions.

As the maximum entropy potential function is relatively well behaved, the problem of early termination may be removed by using suitably small step sizes. However, it was desired to implement an algorithm which may be

faster and more reliable than the univariate search. Research into computational optimisation methods (e.g. Hooke and Jeeves, 1961, Fletcher and Reeves, 1964, and Davidon, 1975) has yielded many concepts useful for an efficient, robust algorithm, such as

(i) **The use of vectors.** Movement in a (well chosen) vector direction can move all ordinates simultaneously at a cost of only one function evaluation, reducing the time taken.

(ii) **Acceleration.** Acceleration along a direction which has been shown to be useful will reduce the number of function evaluations required to reach a certain position.

(iii) **Ridge-following.** If the vector direction is chosen to match the local shape of the potential function, ridges can be followed, reducing the possibility of premature termination.

(iv) **Interpolation.** As the maximum of a quadratic or cubic function can be calculated simply, given the coefficients of the equation, the use of interpolation can allow single steps to be very accurate.

(v) **Quadratic convergence.** Sufficiently close to the maximum, any function may be well approximated by a quadratic, thus algorithms which use the properties of the quadratic function may converge very quickly, once close to the maximum. This property of *quadratic convergence* was not deemed particularly important, as Wernecke and d'Addario (1977) attempted the conjugate gradient algorithm (a quadratically-convergent method) and found that the region of superlinear convergence was too small to give a demonstrable advantage.

It was decided that, for such a complex algorithm, no development should be performed: One of the many algorithms already developed would

be assessed. Attempting to remain true to initial goals (simple algorithm which does not rely heavily on properties of the convolution model), the method selected was that developed by Davies, Swann, and Campey (Swann, 1969). The method uses *N* orthonormal vector search directions, which are frequently realigned to follow the local geometry. The method uses both acceleration and a quadratic approximation to increase the speed of termination. It does not, however, have properties of quadratic convergence.

5.3.1. The Linear search.

The linear search at the heart of the process is an accelerating search, as follows. Given a direction $\boldsymbol{D}$, a step size is defined and a single step taken. If this is an improvement the step size is doubled. This process is repeated until an attempt fails. At this point the mid-point between the final and previous positions is tested, giving four equally spaced points in the vector space. The worst position is discarded (it must be one of the extremities) and the remaining three are used to define a quadratic approximation to the potential function. As the maximum has been bracketed by the accelerating search, the maximum of the quadratic function can be found analytically. Given three positions $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$, with separation S, and three corresponding evaluations of the objective function V_1, V_2, and V_3, the maximum of the quadratic is at

$$\boldsymbol{x}_{\max} = \boldsymbol{x}_2 + \frac{S(V_1 - V_3)}{2(V_1 - 2V_2 + V_3)} D \qquad (5.47)$$

The process is illustrated schematically in figure 5.17.

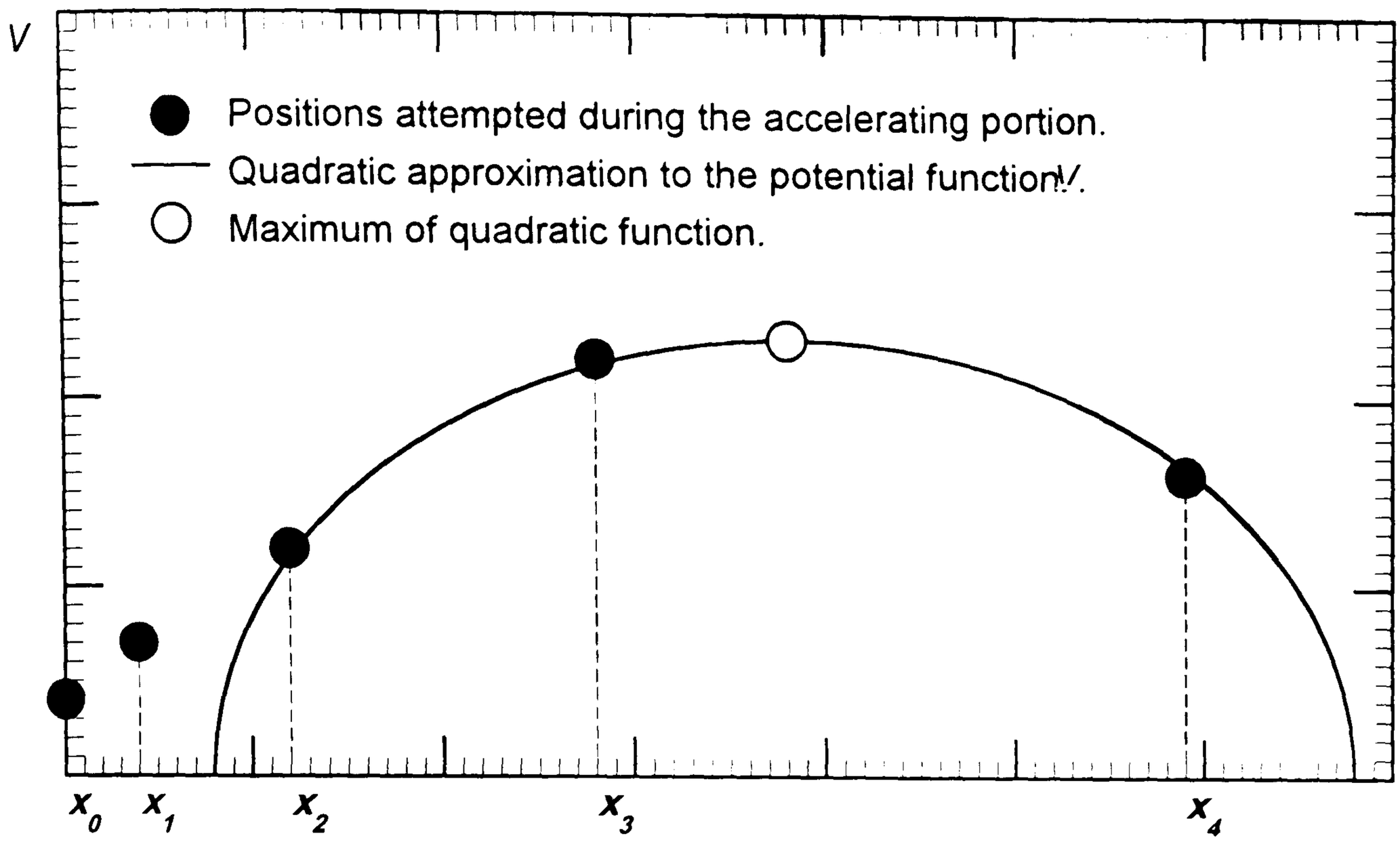


Figure 5.17. Schematic diagram of the linear search from the Davies, Swann, and Campey algorithm.

5.3.2. The search directions.

The vector directions $\boldsymbol{D}_j$ $j=1..N$, are initialised to the co-ordinate directions, thus the first search is univariate. A linear search is made in each direction, and the progress made in each direction, d_j, is calculated. After these N linear searches, the resultant direction and total progress this pass are calculated.

If the total progress made this pass is less than the step size used, it is assumed that the step size is too large. The step size is reduced (divided by 10) and the process repeats without adjusting the search directions.

Otherwise, it is assumed that the resultant direction will be a good direction to continue the search. The step size is kept, and all N vector directions are replaced with N new directions, calculated as follows. New vectors A_k are defined, such that

$$\boldsymbol{A}_k = \sum_{j=k}^{j=N} d_j \boldsymbol{D}_j \qquad k=1..N \qquad (5.48)$$

Thus vector $\boldsymbol{A}_0$ is the sum of all progress made in the pass, $\boldsymbol{A}_1$ is all progress made in directions $\boldsymbol{D}_1$ to $\boldsymbol{D}_N$, and so on. New search directions $\boldsymbol{D}_j^{'}$ are now defined using the Gramm-Schmidt process.

$$\boldsymbol{B}_1 = \boldsymbol{A}_1 \qquad (5.49a)$$

$$\boldsymbol{D}_1^{'} = \frac{\boldsymbol{B}_1}{\|\boldsymbol{B}_1\|} \qquad (5.49b)$$

$$\boldsymbol{B}_2 = \boldsymbol{A}_2 - \left(\boldsymbol{A}_1 . \boldsymbol{D}_1^{'}\right)\boldsymbol{D}_1^{'} \qquad (5.49c)$$

$$\boldsymbol{D}_2^{'} = \frac{\boldsymbol{B}_2}{\|\boldsymbol{B}_2\|} \qquad (5.49d)$$

The new direction $\boldsymbol{D}_1^{'}$ is the resultant direction as required, the other search directions have been adjusted to maintain orthogonality.

A possible problem exists in the orthonormalisation procedure. For any direction which failed to make any progress, $d_j = 0$, and the vector direction would be 'lost'. This loss of orthogonality is avoided by performing the orthogonalisation only on those directions for which $d_j > 0$.

5.3.3. Determining α.

The Davies, Swann and Campey method was tried as a replacement for the univariate search. The method used to determine α was as before: trial-and-error increasing from zero, and using quadratic interpolation.

6.0. Experimental

A number of separate experiments have been performed during this research. The experimental methods have been grouped into this chapter; the results are displayed and discussed in chapter seven.

In this chapter we first introduce the quadrupole SIMS instrument on which most of the depth profiles were taken. The methods and samples used to characterise the noise are then described. Following this are the methods used to reduce the effects of noise on the measurement of the response function. We then introduce the simulated data used to assess the performance of the three algorithms described in chapter 5.0, and show the simulated data sets selected to asses the usefulness of the MaxEnt quantification method on the following commonly measured parameters: the Full Width at Half Maximum (FWHM), the rise and decay inverse slopes, and the peak concentration, and demonstrate the limiting reconstruction case - reconstructing the profile of a true (simulated) delta layer. We then detail three methods attempted to improve the performance of the method on the background signal. Finally we describe some tests performed to check the validity of the method on real SIMS data and the real samples quantified by MaxEnt.

6.1. The EVA2000 Quadrupole SIMS Instrument.

The investigation of the noise on a SIMS depth profile, and the measurement of response functions and data for reconstruction was performed on the in-house EVA2000 SIMS instrument. The instrument was built by Dowsett some thirteen years ago (Dowsett and Parker, 1983,

Dowsett, Heal, Fox, and Parker, 1985), but has since been substantially modified. The instrument will be described in its current configuration.

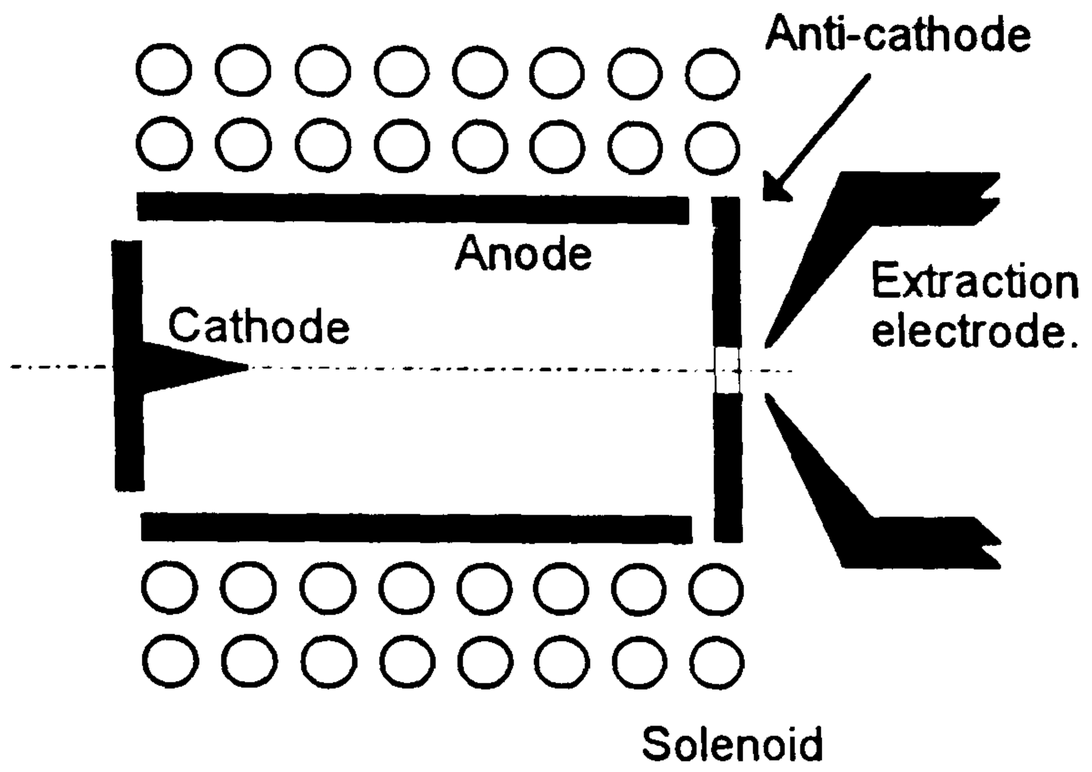


Figure 6.1. Schematic diagram of the Penning ion source.

The primary ions are generated in a Penning discharge source, recently re-designed (Cooke, 1992) to simplify the maintenance procedures. A pressure of $3x10^{-6}$ mBar of high purity (99.998%) O_2 molecules is maintained in a chamber manufactured from high-purity aluminium components (Dowsett and Parker, 1983). This gas is ionised predominantly by electron impact, using electrons emitted from the cold cathode by ion impact; an axial magnetic field is used to increase the electrons' path length.

The primary ions are extracted through a 0.7 mm hole in the anticathode by an electrode which is a few hundred volts negative with respect to the plasma. This forms part of a telefocus immersion lens system in which the bulk of the accelerating field is applied across a gap farther down the ion column.

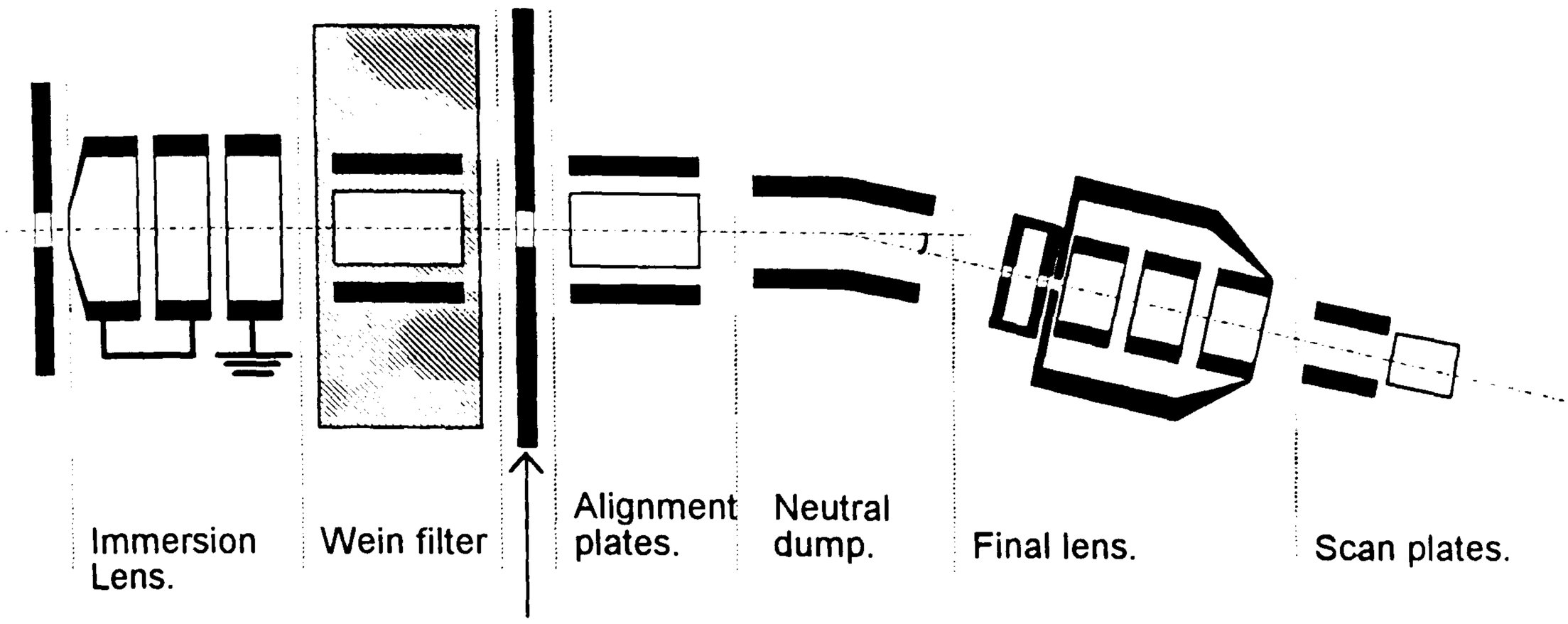


Figure 6.2. Schematic diagram of the Primary ion column.

The ions from the plasma meniscus are focused by the immersion lens through a Wein filter, where crossed electric and magnetic fields perform mass/charge selection, into a waist between the pressure step and the inlet aperture to the final lens. (The distance between this waist and the final lens controls the angle of divergence into the lens, thus limiting the spherical aberration). A pair of alignment plates in the Wein filter allow the beam to pass through the pressure-step aperture into the lower column, maintained at the UHV pressure of the main chamber (10^{-8} mBar). After alignment plates (which are not currently necessary) a 4° bend and an electrostatic field remove any neutral beam component. The ions then pass into the final immersion lens to be focused onto the sample. Immediately after the lens are two sets of plates which perform the scanning of the beam. To give a uniform ion flux across the crater bottom, the raster electronics use high-linearity 14-bit DACs. Angular alignment of the raster pattern may be performed by rotating the entire final lens assembly.

Samples are mounted on a stainless steel sample holder, fixed using conductive paint. The sample holder may be electrically biased to optimise

secondary ion emission. To maintain UHV conditions in the chamber, samples are loaded and removed via a load-lock.

Secondary ions are accelerated into an energy filter (Wittmark, Dowsett, and Clegg, 1982) by a potential difference of approximately 200 V for positive secondary ions. The energy-filtered ions then pass through an electrostatic iris (Dowsett and Parker, 1983), which may be used to attenuate the signal (to protect the detector). The ions pass into the quadrupole mass filter, which uses 19 mm diameter rods giving a total mass range of 200 amu. Electrostatic deflection plates then bend the mass filtered ions into the detector, a Channeltron operating in pulse-counting mode.

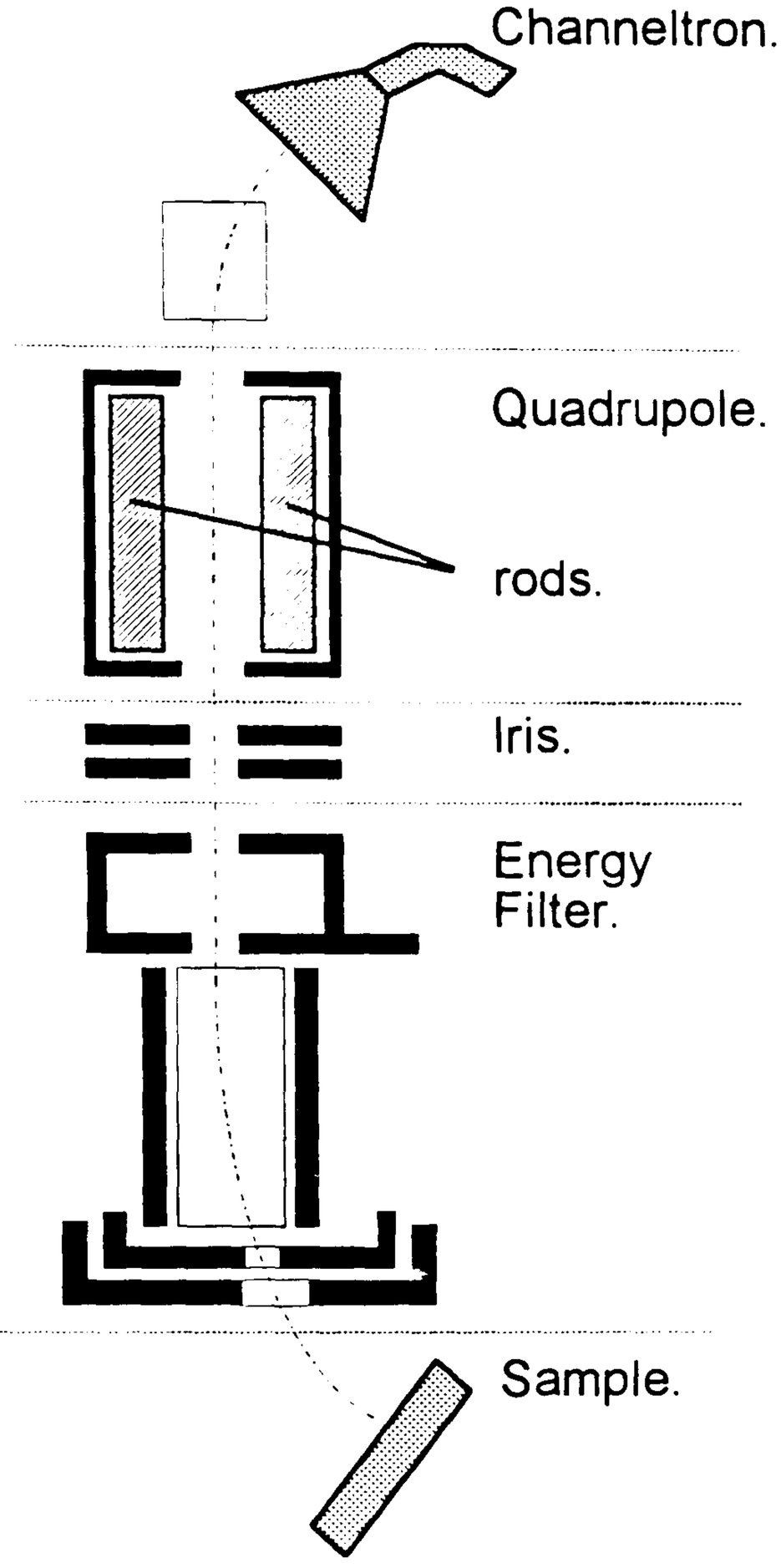


Figure 6.3. Schematic diagram of Secondary ion column.

Although the instrument was originally designed to be operated manually, much of the instrument is now computer controlled; purpose-built interface electronics allows control by a Research Machines (RM) Nimbus 380Z. The control programs are written in RM-BASIC with machine code extensions (Dowsett, Heal, Fox, and Parker, 1985). The depth profile data are stored in memory during the profile, then saved to a 5.5" disk for later

transfer onto an IBM-PC compatible disk. Modifications to the software by the author allow the data to be sent to the PC as the experiment continues, removing the laborious task of transferring the files and, more importantly, removing the memory-based limitation of three thousand data points per depth profile.

6.2. The characterisation of the noise on a profile.

In section 4.4 it was argued that the noise on a SIMS depth profile would have components of both Poisson and Gaussian distributions, but might be adequately described by the Gaussian distribution as required by the MaxEnt method as described in Chapter 3.0. Before the reconstruction method was tested, this distribution was investigated and characterised. As previously stated experimental method only is grouped here, with results in section 7.1. For this investigation we shall describe first the types of sample selected as suitable, the methods used to characterise the distributions, and then the individual experiments.

6.2.1. Suitable samples for the investigation.

The standard method for characterising noise during an experiment is to make repeated measurements and examine the distribution of the results. Provided enough measurements are taken, in the absence of systematic errors, the mean will be a good approximation to the true value and the measured values will form some distribution about this mean. For SIMS depth profiles it is theoretically possible to compare equivalent frames in many different profiles of the same sample. However, this would require a very large amount of lateral homogeneity in the sample, and a degree of instrumental stability that is seldom found in SIMS. It was decided that a

method more likely to succeed was the use of a sample such that each frame is, effectively, a repeat of the last. We might then examine the variation throughout the profile to obtain the mean and the distribution.

The constant concentrations required for this type of experiment are available in elementally pure samples. However, analysis conditions are often modified to attenuate the high intensity signal from matrix concentrations, to prolong the lifetime of the detector. This could lead to a different balance in the causes of noise on the profile, thus it was felt that such samples are of limited use; the noise should be measured from dilute features where possible. Assuming that a suitably dilute sample could be found, it was also desired to investigate the variance of the distribution as a function of signal intensity, due to the Poisson noise component. It was also considered prudent to investigate the possibility that the characteristics of the noise vary with experimental conditions or sample constituents, where possible. For practical reasons it was desired to investigate all these possibilities from a single sample. It was decided that a suitable sample would be a staircase structure as grown by MBE for MBE calibration purposes. These structures contain (ideally) a number of different regions of constant concentration. The regions could be investigated separately to give the distribution at each mean concentration level.

6.2.2. Methods for obtaining the distribution of the noise.

6.2.2.1. Constant signal approximation.

If the instrument is perfectly stable the steady-state signal from a region of constant concentration will be constant except for the noise. Within this approximation the noise distribution may be obtained by plotting a

histogram of the number of occurrences of individual intensity values. For samples where the range of values is large the number of occurrences of each value may be small. Here the form of the histogram may be made more obvious by grouping the intensity values into 'buckets' spanning a range of intensity values.

From the profiles of the samples (to be described) the mean and standard deviation of the values was calculated, and a Gaussian distribution with these parameters overlaid. For the low intensity signal, where the distributions have small mean values (less than about 10 ions frame^{-1}), Gaussian and Poisson differ significantly, thus a Poisson distribution was fitted to the histogram (using SigmaPlot 5.0, Jandel Scientific, California) and overlaid. Visual inspection was used to determine whether either superposed distribution was a good approximation to the histogram.

6.2.2.2. Linear variations allowed.

Experience with the above method indicated that for some profiles the approximation is poor: slow variations of the matrix signal may be seen. The most likely causes were expected to be unmeasured variations of primary ion current, or small amounts of charging during the analysis. As the EVA2000 instrument is not capable of measuring either of these effects, it was attempted to allow for these variations, wherever they occurred, by fitting a linear relation to the data

$$Y = mf + c \tag{6.1}$$

and subtracting to yield a histogram of the residuals. This histogram was then overlaid with Gaussian and Poisson distributions as above. This method, however, is only suitable for investigating the distribution of the noise; it is difficult to realistically calculate a mean value in this case.

6.2.2.3. Higher order fitting.

In cases where a linear fit did not adequately follow the trends of the signal, a third method was considered using a quadratic fit for the calculation of the residuals. An equation of the form

$$Y = af^2 + bf + c \qquad (6.2)$$

was thus fitted to the measured data and then subtracted to obtain the residuals. Again, with this method it was not possible to investigate a variation of standard deviation with the mean signal intensity.

6.2.3. A preliminary investigation.

As a preliminary investigation it was attempted to characterise the noise on the matrix channel, and that on the low-intensity signal (henceforth termed the *background* noise) separately.

6.2.3.1. The matrix channel.

Although it was expected that matrix signals might not be representative of the noise on a dilute impurity profile, a preliminary investigation was performed using the silicon channel from the available depth profiles of a boron in silicon implanted standard (10^{15} atoms cm^{-2}, implanted at 25 keV). For this investigation the possibility of a primary ion energy dependence was ignored; profiles from 1.5 to 9 keV ion^{-1} were used. These profiles had been previously measured in the EVA2000 instrument by Mr. R.D. Barlow, using normal incidence $^{16}O_2^+$ ions. Other instrumental conditions were standard: frame times of 10 or 5 seconds at primary ion energies up to and above 6 keV respectively; the rastered area was nominally 400 μm along each side, digitally gated such that the measured signal comes only from the central 125 μm square. The onset of equilibrium

for these samples was determined by eye, and the distribution of the noise investigated using all three methods.

6.2.3.2. The background signal.

A preliminary investigation of the background noise was performed using two samples. Efficient use of instrument time includes profiling only as far as is necessary, thus depth profiles of wide features leave few data points from which to investigate the background distribution.

Accordingly, an investigation was performed using depth profiles of boron delta layers in silicon. The analysis was performed in the EVA2000 instrument, using 4 keV ion^{-1} $^{16}O_2^+$ at normal incidence. Primary ion current was 150 nA, raster size was 410 μm, digitally gated to the central 125 μm, frame time 10 seconds.

A second preliminary analysis was also made using data from thin silicon features in gallium arsenide provided by Prof. M. Maier. These GaAs(Si) samples were bombarded using 6 and 15 keV ion^{-1} Cs^+ ions at 45°, 50 nA. Craters were 500 μm wide, gated to the central 30%, with a frame time of 1 sec. The $AsSi^-$ channel was detected.

6.2.4. Investigation using a staircase sample.

Following the preliminary investigations, a more complete experiment was performed using a boron in silicon staircase sample. The sample was grown by MBE in the modified VG Semicon V90 instrument at Warwick; sample identification number 17/3. It contained three mean doping levels: nominally 10^{17}, 10^{18}, and 10^{19} atoms cm^{-3}, of widths 350, 200, and 350 nm respectively. It was profiled twice in EVA2000 using normal incidence $^{16}O_2^+$

at 4 keV ion^{-1}, primary ion current 150 nA, frame time 10 seconds, rastered area 410 μm, digitally gated to the central 125 μm. One of the depth profiles, calibrated by crater depth measurement and comparison with an implanted standard, is shown in figure 6.4.

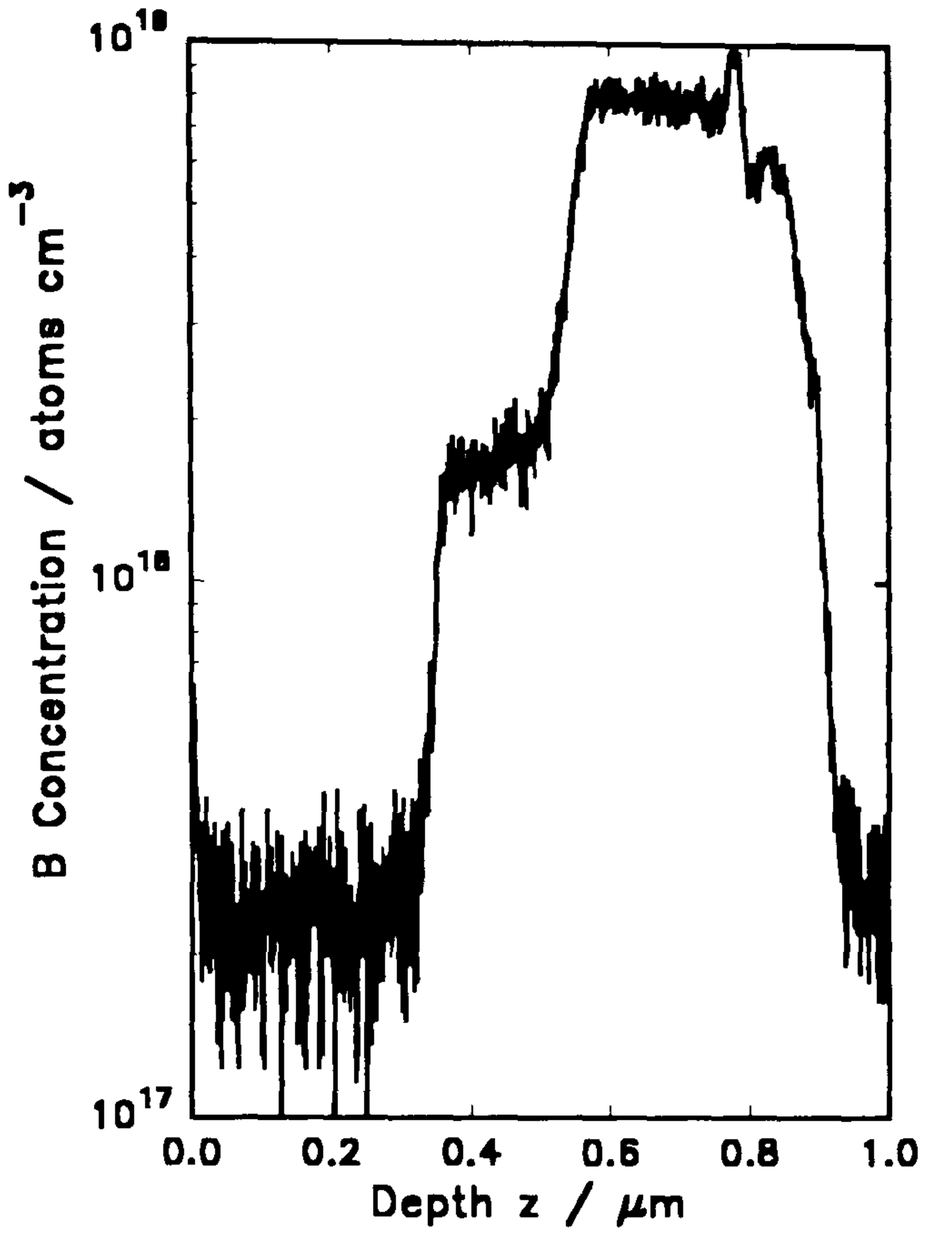


Figure 6.4. The Si(B) staircase 17/3, profiled at 4 keV ion^{-1}.

Thirteen point running-average smoothing (using the quantification software ProWin, described in chapter 8) was used to determine the extent of each constant-concentration 'step' and the transition regions. The (unsmoothed) data were then divided into separate regions and the transition regions discarded. The MBE growth effect at 0.8 microns (figure 6.4) was also discarded. For these samples only method 6.2.1.1 was used to determine the noise distribution.

6.2.5. Investigation of the energy dependence.

To investigate the energy dependence of noise on the features, another Si(B) staircase sample was used: concentrations varying between

10^{16} and 10^{19} atoms cm^{-3} over a depth of 1.2 μm. The sample (MBE identification number 33/02) was grown by MBE at Warwick, growth temperature 600°C. The sample was profiled in EVA2000 using normally incident $^{16}O_2^+$ primary ions at energies of 4, 6, and 8 keV ion^{-1}. Experimental conditions are shown in table 6.1.

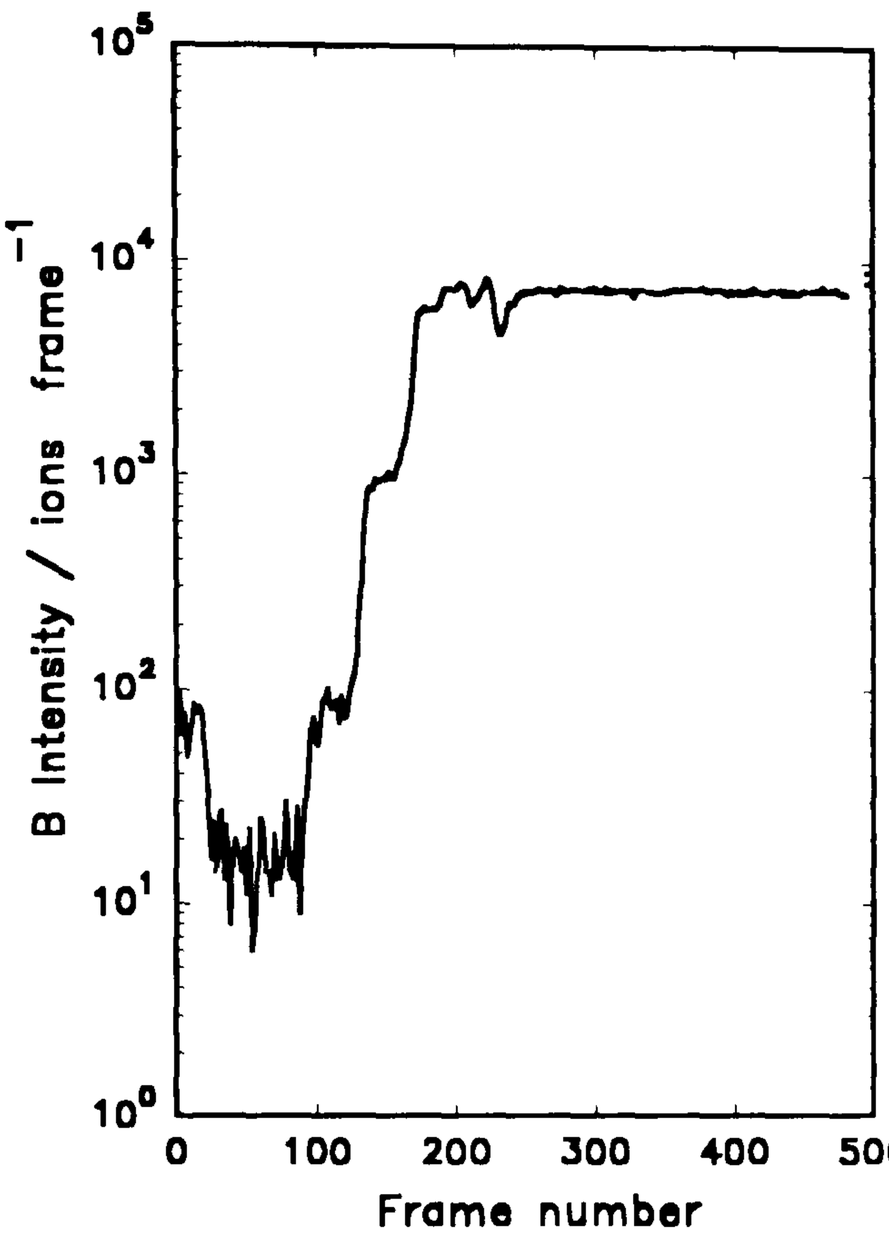


Figure 6.5. The Si(B) staircase 33/02, at 8 keV ion^{-1}.

Primary Ion Energy / keV	Primary Ion current / nA	Frame Time / s	Rastered Area / μm
4	300	10	410
6	600	10	402
8	800	10	404

Table 6.1. Analysis conditions for the Si(B) staircase sample 33/02.

As before, smoothing was used to determine the extent of the steps and transition regions; the data were sectioned and the MBE 'bump' visible in figure 6.5 discarded. Again, the method of section 6.2.2.1 was used to characterise the noise.

6.2.6. Investigation of the elemental dependence.

Finally, to test for a dependence on the element, a staircase in another material was sought. The only one available was in the matrix regime: a silicon/germanium staircase. The germanium concentrations vary from nominally 70% down to 15% over 1.9 μm, the layers deposited at 500°C.

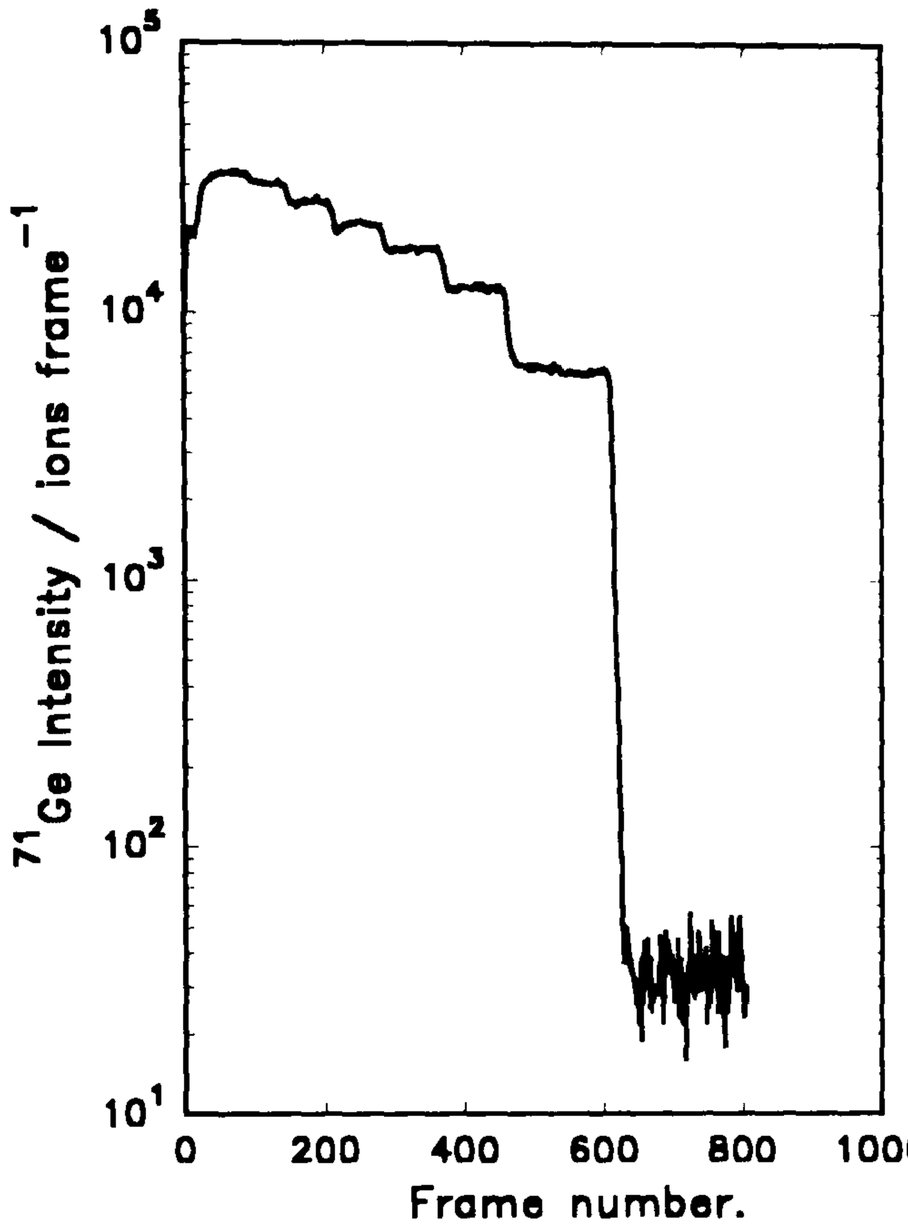


Figure 6.6. The SiGe staircase 24/16, profiled at 6 keV ion^{-1}.

The sample was analysed in EVA2000, using $^{16}O_2^+$ ions at normal incidence; impact energies of 4 and 6 keV ion^{-1} were used. To protect the detector from the high intensity signal, the secondary ion emission was retarded by applying a negative voltage to the sample stage, in this case -20 V for the detection of silicon and -30 V for germanium. Frame time was 10 seconds throughout and raster size 410 μm with standard (125 μm) gating. One of the raw 6 keV profiles is shown in figure 6.6. The profiles were sectioned as before and method 6.2.2.1 was used to determine the variance of the noise.

6.3. Reducing the effect of noise on the response.

The noise on a SIMS depth profile is not a part of the response of the instrument, but occurs outside the convolution (section 4.4). The measurements of delta functions are subject to the same additive noise as any other profile. To extract the response as accurately as possible from a measured depth profile this noise must be removed. Fortunately, both theoretical predictions and pragmatism lead to a method by which the effect of the noise may be reduced to below other experimental errors.

6.3.1. The form of the response.

The profile measured from a delta layer is very often a characteristic shape: far from the peak the form is double-exponential, closer to the peak the form is more or less curved, depending on the sample. Theoretical treatments of the atomic motions have predicted such forms (Littmark and Hofer, 1980), and the form may be fitted to almost any delta profile (Dowsett, Rowlands, Allen and Barlow, in press).

6.3.2. Extrapolation.

As the exponential behaviour extends over two to four orders either side of the peak, it is not unreasonable to assume that the true response continues to fall exponentially either side of the peak and perform some extrapolation. Even were this not true, the ability to measure the instrumental response over potentially five orders of magnitude means that the errors introduced by *any* assumptions about the response below the background signal will be proportionately tiny. Over a reasonable profile (four orders), the

errors due to the assumption of any form below the fourth order amount to 0.01%, far below all other experimental errors.

6.3.2.1. Best fit by eye.

On a logarithmic plot, as is commonly used in SIMS, exponential slopes are straight lines. It is straight-forward to select the two linear regions of the graph, draw a line of best fit by eye, determine the slope, and extrapolate. To automate this process slightly the replacement SIMS quantification software ProWin (described in chapter 8.0) was adjusted: the user may use cursors to define the line of best fit, and the computer performs the extrapolation.

6.3.2.2. Computed least squares best fit.

As the main reason for the use of maximum entropy in the reconstruction process is to reduce the possibility of bias in the results (section 3.0), it was considered reasonable to attempt to exclude bias from the extrapolation process. To achieve this, the line of best fit might be determined computationally using the theory of least squares (Heading, 1970). Given that we wish to determine the best straight line *f(z)*

$$f = mz + c \tag{6.3}$$

to the data points

$$(y = \log(C), z) \tag{6.4}$$

we may calculate the sum of squares of the deviations of *f* from *y*,

$$S = \sum (y - mx - c)^2 \tag{6.5}$$

and, by setting the partial derivatives to zero obtain the value of *m* as

$$m = \frac{\sum_i Z_i Y_i}{\sum_j Z_j^2} \tag{6.6}$$

where Y_j and Z_j are the deviations from the mean values

$$Z_j = z_j - \bar{z}, \text{ and} \tag{6.7a}$$

$$Y_j = y_j - \bar{y} \tag{6.7b}$$

The value of c may be determined from equation 6.3 as the line of best fit passes through $\bar{z}, \bar{y}$.

To automate this process, ProWin was adjusted such that the user specifies the straight region and the computer calculates the regression line and performs the extrapolation. As the sum of squared residuals is simple to calculate, it is possible for the computer to test the linearity of the users suggested region, and to search for better regions nearby. Such a method should be totally free from user bias. Unfortunately, as the 'best' region as determined by the simple minimum sum-of-squares is likely to be the shortest possible (two data points), it was decided unwise to allow the computer total control; it merely suggests the 'better' region, and the user must decide whether or not it should be used. It was felt that the time required to devise an algorithm which attempted to find the minimum sum-of-squares over the longest possible line was prohibitive.

6.3.3. Fitting the form to the entire profile.

The use of simple assumed forms for the response of the instrument, such as Gaussian or double-exponential, has been avoided during this research, as any assumption of the form involves user bias. However, after work in this thesis was completed a more complex function was developed which has shown great promise. The function (Dowsett, Rowlands, Allen, and Barlow, in press) is the convolution of a Gaussian function with a double exponential function. This convolution, performed analytically, allows very fast fitting of the form to the profile of a delta layer. Thus far the form has

fitted profiles of both boron in silicon and silicon in gallium arsenide, which are of greatly different form, over many orders of magnitude. The three shape parameters obtained correspond to the standard deviation of the Gaussian component and the two *remote* inverse slopes. These parameters were found to have a much better defined energy dependence that those previously extracted from delta profiles. The fourth parameter defines the position of the response with respect to the origin.

6.4. Appraisal of the optimisation algorithms.

The initial histogram algorithm had demonstrated the feasibility of maximum entropy in SIMS using small simulated depth profiles and an analytic response (section 5.1). The first priority of this research was to adjust the algorithm to extend this feasibility to realistic data sets. A full SIMS depth profile from the EVA2000 instrument may consist of three thousand data points. The profile of the impurity would be interlaced with a reference channel, leaving fifteen hundred as the maximum data-set. This was initially far beyond the abilities of the program. It was decided that a realistic figure for the minimum number of data points should be 100; under typical operating conditions this corresponds to about 80 nm of eroded silicon, which is large enough to contain the not-too-thin features expected to benefit most from this method. The speed of the algorithms was thus to be investigated with regard to two criteria.

(i) During the development stage the speed required to converge for a small test data set was used to rank the algorithms and any proposed modifications. As the major computational cost of all the algorithms is the calculation of the convolution integral, the number of convolutions required to converge serves as a useful measure, independent of the

computer used to perform the investigation. After this initial development stage the most efficient variant of each algorithm was retained.

(ii) As the size of the hyperspace to be searched is proportional to the N^{th} power of some constant x, it is common for the time taken by an algorithm increase non-linearly with N, or even as x^N. The behaviour of the algorithm as N was increased was thus assessed, to select the algorithm most likely to be efficient for large data sets.

Accordingly, two sets of simulated data were constructed using a measured response function for the boron-in-silicon system. The method was as follows:

(i) The response of the EVA2000 instrument was measured.

(ii) A 'true' concentration profile for this system was assumed.

(iii) The assumed concentration profile was convolved with the response.

(iv) Gaussian noise was added to the simulated profile

(v) The results were then rounded to the nearest integer value, to remove the possibility of the extra information yielding abnormally good results.

6.4.1. Speed on small data sets.

To generate a standard piece of test data with which to compare the performance of the algorithms and variants, the response function for the EVA2000 instrument was measured. The sample was an MBE-grown boron in silicon delta layer (sample identification number 10/33), $3x10^{14}$ atoms cm^{-2}, 51 nm below the surface, previously characterised using XTEM and SIMS (Powell et.al, 1991b). The response was generated as described in Section 4.2.4.1 (constant erosion rate, peak placed to match expected differential

shift), using the extrapolation method of section 6.3.2. It is shown in figure 6.7a.

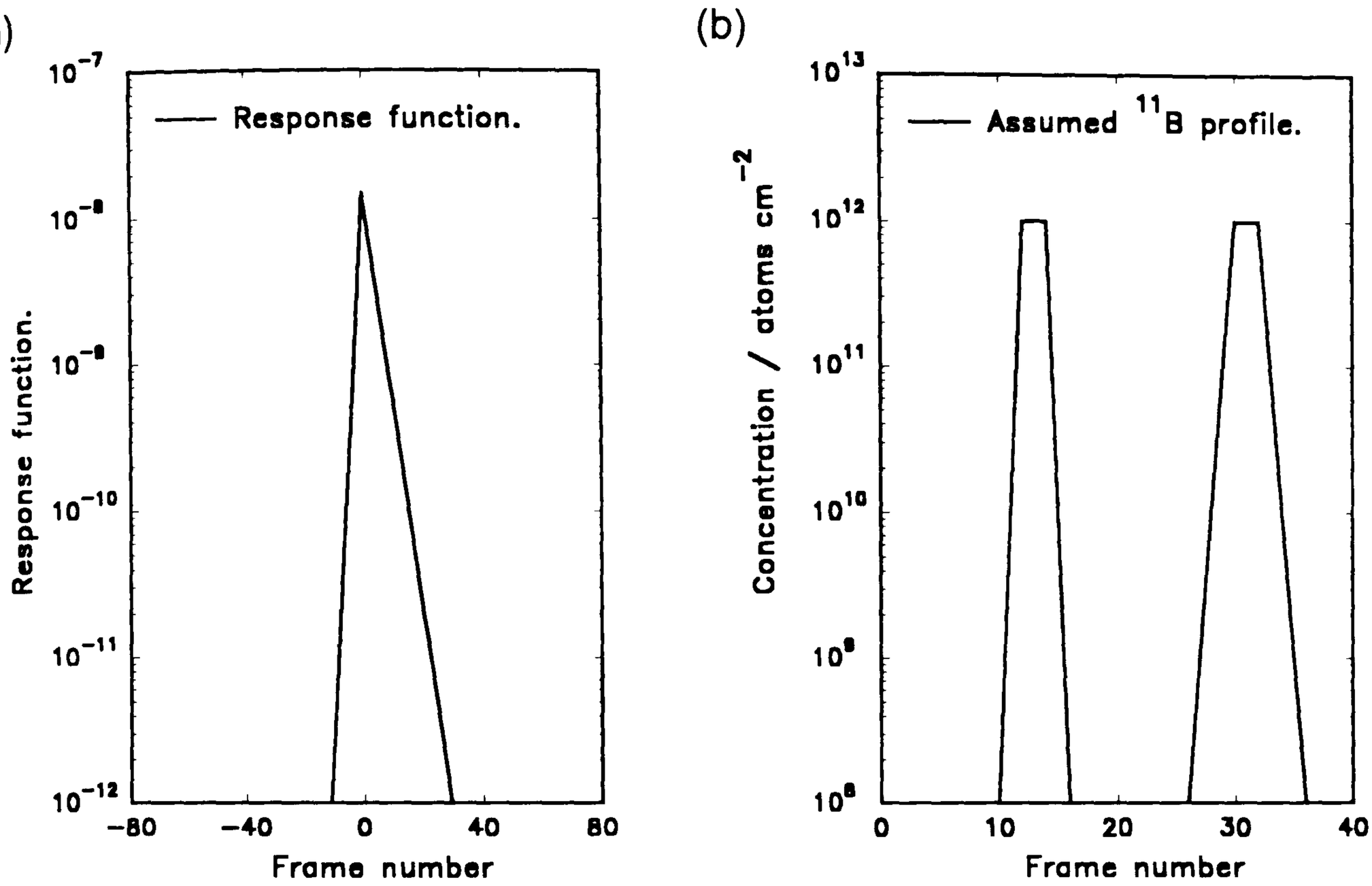


Figure 6.7. Speed Data Set 1: (a) the response function, and (b) the assumed concentration profile.

The assumed profile is shown in Figure 6.7b: two thin but slightly different layers. The noise added was Gaussian, with standard deviation σ given by the results of the preliminary noise study:

$$\sigma = 0.0168Y + 86.6 \qquad (6.8)$$

where Y is the value of the simulated SIMS profile in that frame. The predicted SIMS depth profile is shown in figure 6.8. It has relatively few data points, allowing quick assessment of each algorithm. The amount of background noise was later found to be overestimated, however, as discussed in section 7.1.

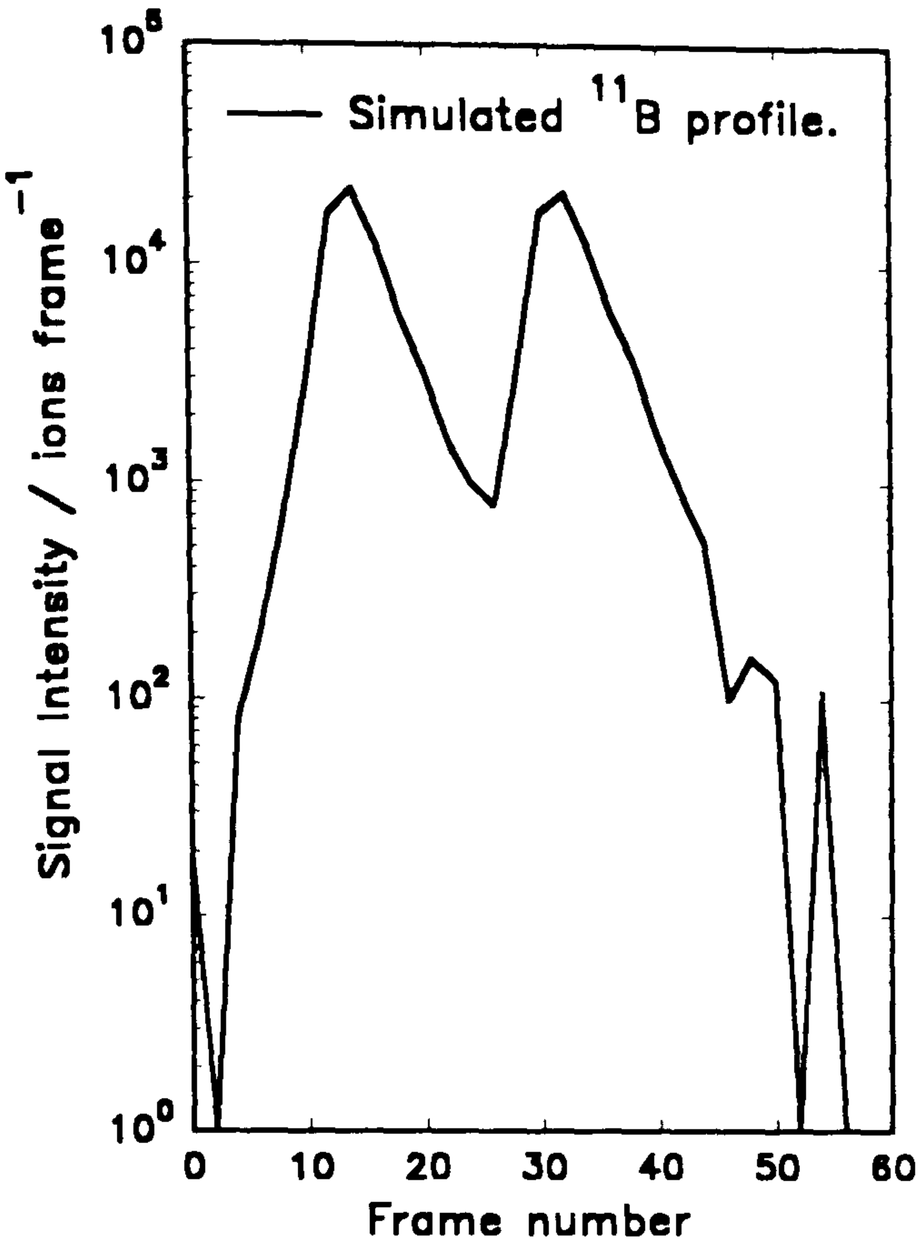


Figure 6.8. Speed Data Set 1: the simulated SIMS profile.

6.4.2. Speed as the number of variables increases.

To test the speed as a function of *N* without artefact, it was attempted to create a set of simulated profiles *of the same feature*. To allow meaningful representation of a distribution for small *N* a simulated profile which contained only a single peak was constructed. Identical concentration distributions were created for *N*=16, 32, 64, and 128 (figure 6.9a). Response functions were generated (again from delta layer 10/33) with a corresponding number of data points (32, 64, 128, and 256) (figure 6.9b). The data were convolved to simulate the profile and noise of standard deviation

$$\sigma = 2\sqrt{Y + 5} \qquad (6.9)$$

added, in accordance with the final results of the noise study. The data were then rounded to the nearest integer. The corresponding simulated depth profiles are shown in figure 6.10.

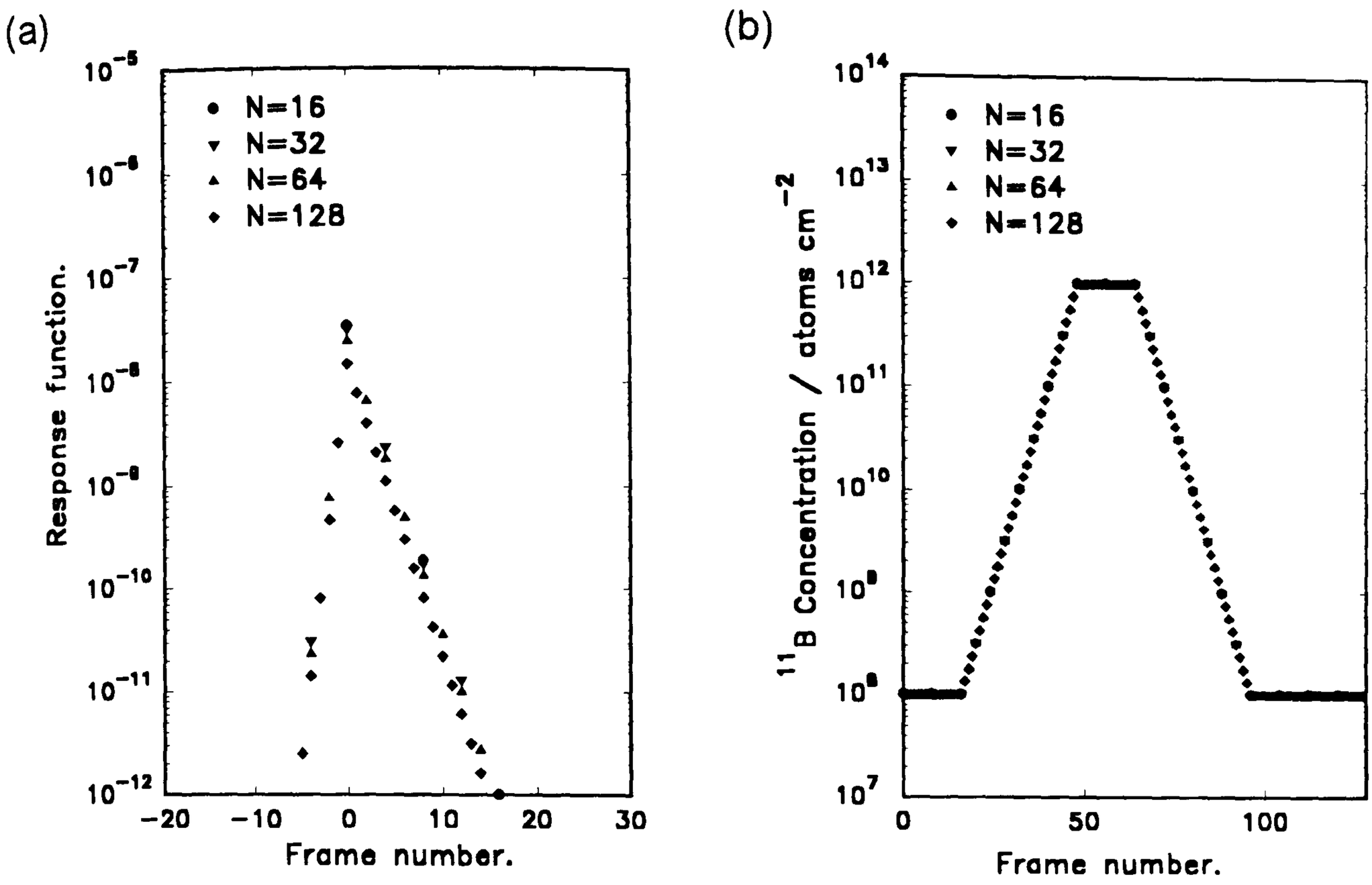


Figure 6.9. Speed Data Set 2: (a) The response functions, and (b) the assumed concentration profile.

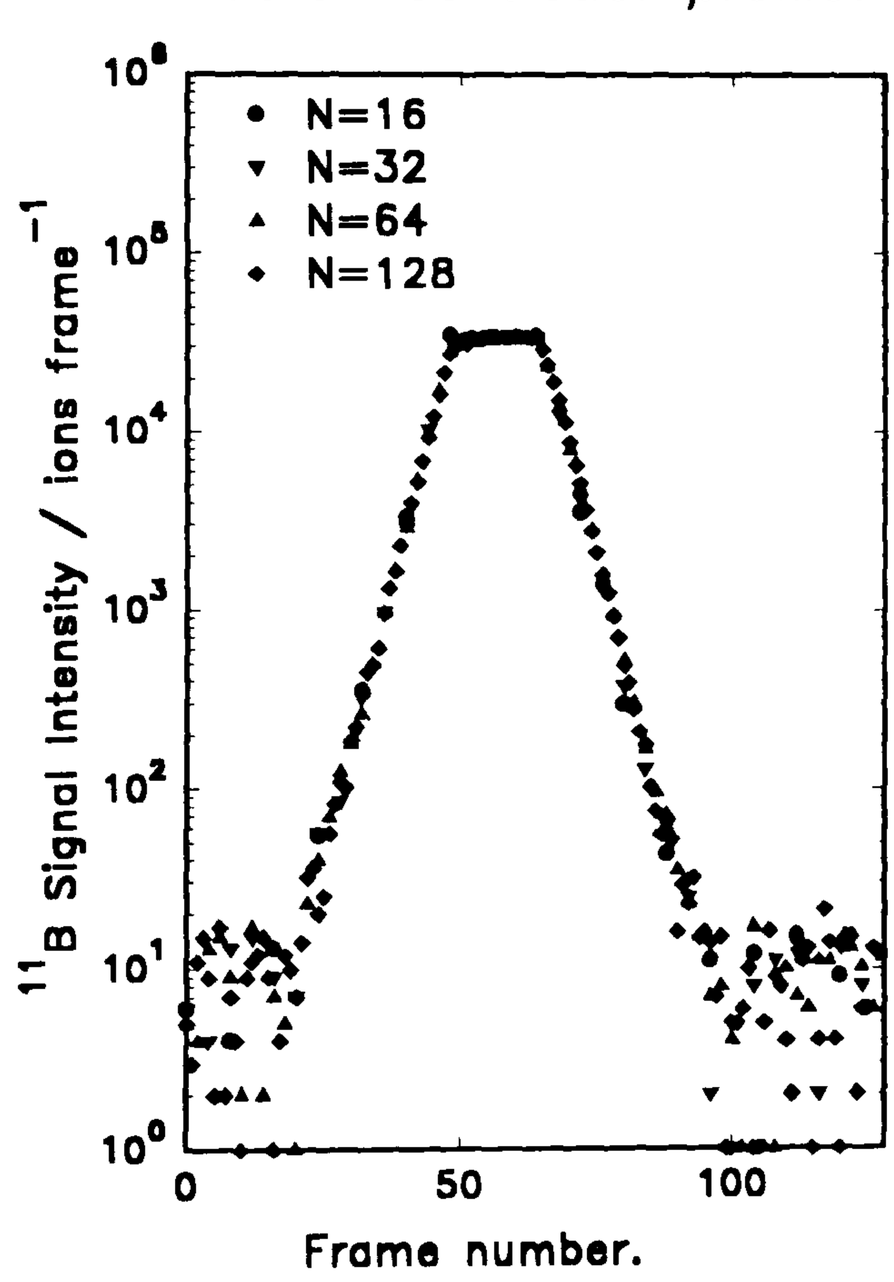


Figure 6.10. Speed Data Set 2: the simulated SIMS profiles.

6.4.3. Data reduction methods.

The development stages yielded algorithms which could realistically reconstruct data sets of one or two hundred data points. It was decided that some method of reducing the size of a profile might be used to allow the reconstruction of larger profiles. A number of methods were considered.

6.4.3.1. The reconstruction of profile segments.

Where only a subset of the profile is of interest, or only a subset is expected to benefit from the use of the MaxEnt quantification method, the profile may be sectioned, and reconstruction performed only on a limited part of the data. If this method is performed by segmenting the raw data, accuracy will be lost at the edges of the region, where a restricted part of the available information is used. This could be avoided by specifying overly large regions of interest, but no simple method could be devised to estimate the varying accuracy of the reconstruction.

It would be better to perform such sectioning by restricting the number of ordinates but performing the calculation of the mismatch over the whole profile. In this way all the available information would be used. This would require some adjustment to the theory, however, as the choice of the parameter α in the potential function depends on N (section 3.3.1). We would now have the case that the mismatch is calculated over N points, but the $M<N$ variables do not span the range and the optimum will be found at some $\chi^2 < N$.

6.4.3.2. Neglecting data points.

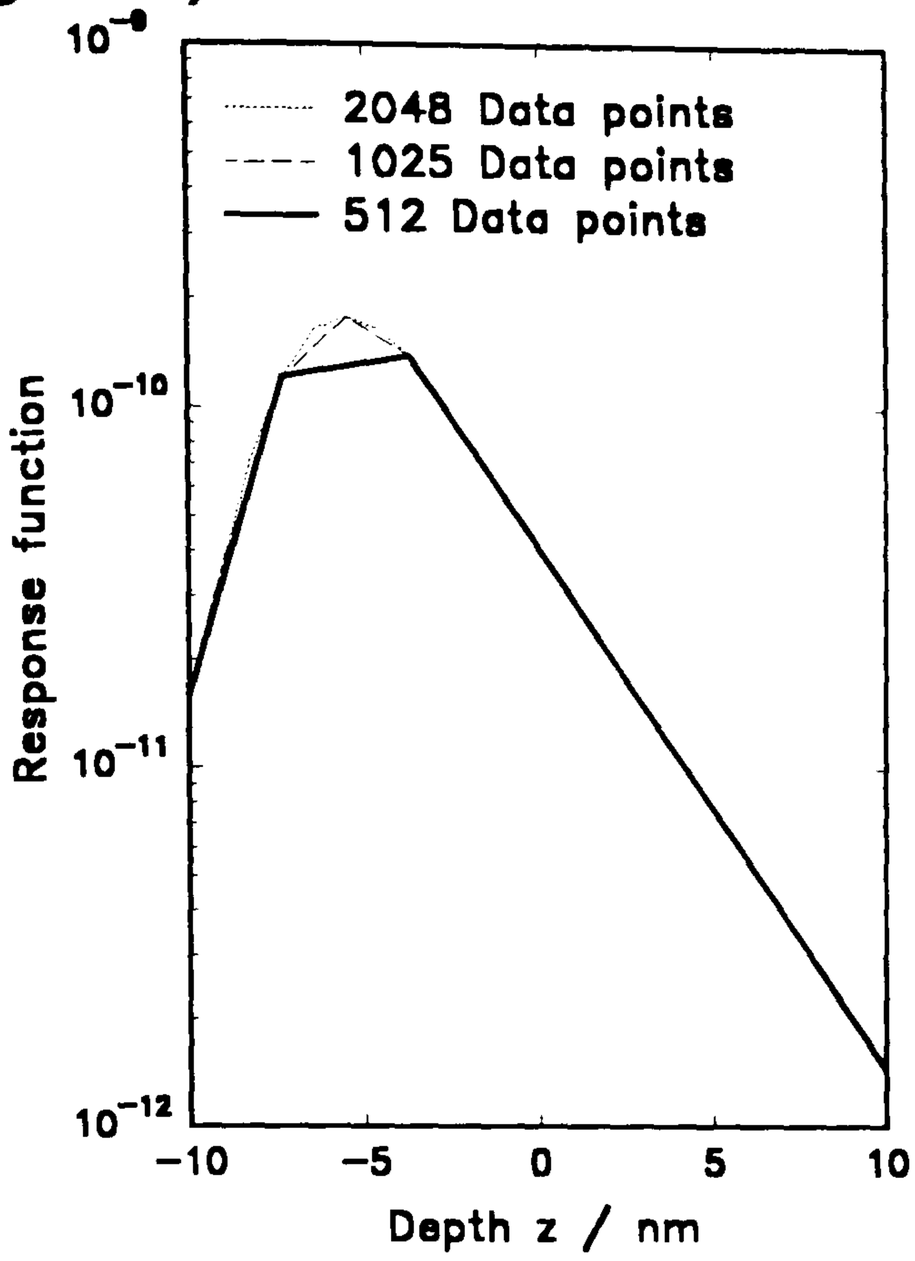


Figure 6.11. The effect of neglecting data points on the response function.

Another brute-force method to reduce the number of variables is simply to neglect alternate data in the depth profile, correspondingly reducing the number of data points in the desired reconstruction. To remain consistent, the number of data points in the response should also be reduced, and a compensating increase in the effective erosion rate must be made. For this method there are no edge regions of unknown accuracy and the value of α such that $\chi^2 = N$ may be legitimately selected. However, the process of removing data points is discarding information: precisely what we are trying to avoid in the quantification. This involves a loss of accuracy in the shape of the response, as shown in figure 6.11. Although sufficient care might reduce the loss of accuracy, it was decided that this method should be avoided.

6.4.3.2. Summing over consecutive data points.

An alternative method devised is to add the value in consecutive frames to form a depth profile which was effectively performed with a longer frame-time. If the data are added in pairs we effectively double the frame time and halve the number of variables. We must again remain consistent and perform an identical process on the response. Again, however, the process is utilising less than all the information. It was eventually decided to reconstruct only profiles where the entire profile could be used, or at worst to neglect a few points of background noise if the reduction in the number of variables would be significant.

6.5. Appraisal of the MaxEnt method.

The quantification of SIMS depth profiles using the MaxEnt method is conceptually more difficult and practically more time consuming than the conventional methods. It was decided to use the convolution model firstly to determine the types of features for which MaxEnt would yield greater accuracy than standard quantification should be identified. Secondly, as the convolution is an information destroying operation, it was attempted to determine how close to the original profiles the MaxEnt reconstruction would be. Thirdly, it was attempted to determine how much better (if at all) the MaxEnt quantification of SIMS depth profiles is than simpler, faster deconvolution methods such as Fourier deconvolution. In order to perform these tests, samples were required for which the 'true' distribution was known: simulated data was again needed.

To quantify the improvement it was decided that a visual comparison of the results might be informative but would also be subjective, while the

calculation of a misfit statistic would be objective but difficult to interpret. Accordingly, depth profiles were simulated for a few carefully selected types of feature, for which the parameterisation would be straightforward.

We shall now describe first the methods used to assess the region of improvement, amount of improvement, and relative performance of the MaxEnt and inverse Fourier methods, then detail the simulated depth profiles used. The results of these experiments can be found in chapter 7.0.

6.5.1. Identification of the expected region of improvement.

To identify the features for which a significant improvement could be obtained using MaxEnt, the parameters extracted from the simulated depth profiles were compared with those from the assumed concentration profile.

6.5.2. Quantifying the improvement gained by reconstruction.

To quantify the improvement gained by using MaxEnt reconstruction and determine how close the reconstructed profile is likely to be to the true profile, the simulated depth profiles were reconstructed using the univariate MaxEnt code. The appropriate parameters, measured from the reconstruction, were compared to those of the simulated depth profile and the assumed distribution.

6.5.3. The improvement gained by Fourier deconvolution.

To compare the results of MaxEnt reconstruction with a simpler, faster deconvolution method, the deconvolution of the simulated depth profiles using division in Fourier space was also performed. The parameters extracted from the deconvolution result were then compared to those of the MaxEnt method.

To meaningfully use the inverse Fourier deconvolution method, some smoothing must be performed. A number of smoothing techniques were briefly assessed for use in these tests. The methods were:

(i) Repeated *n*-point smoothing (until all negative values were removed from the solution).

(ii) Multiplication of the (complex) deconvolved spectra in frequency space by the 'Blackmann' window suggested by (Tagle, Martinez Sáez, Rojo, and Salmerón, 1978, and Chornik, Sopizet, and LeGressus, 1987):

$$W_b(k) = 0.42 + 0.5\cos\left(\frac{\pi k}{K}\right) + 0.08\cos\left(\frac{2\pi k}{K}\right) \qquad k < K \qquad (6.10)$$

$$W_b(k) = 0 \qquad k \geq K$$

where the cut-off value *K* is selected by the user to remove the effects of noise amplification, such as negative concentrations, or 'ringing'.

(iii) Multiplication by one of the many frequency windows suggested by Nuttall (1981), selected for the swift fall-off of its sidelobes:

$$W_n(k) = \frac{1}{32}\left[10 + 15\cos\left(\frac{\pi k}{K}\right) + 6\cos\left(\frac{2\pi k}{K}\right) + \cos\left(\frac{3\pi k}{K}\right)\right] k < K \quad (6.11)$$

$$W_n(k) = 0 \qquad k \geq K$$

where the cut-off value *K* is again selected by the user.

The assessment indicated that both spectral windows were significantly better at removing noise without causing excessive blurring than the repeated *n*-point smoothing. In view of the behaviour of the sidelobes of the windows, the window suggested by Nuttall (1981) was selected for use in these tests.

6.5.4. The simulated depth profiles.

6.5.4.1. Parameters investigated.

The features were selected to be fully described by a single parameter. The parameters selected were those commonly used in SIMS depth profiling (Dowsett and Barlow, in press):

(i) The full width at half the maximum value (FWHM) of a feature.

(ii) The rise and decay slopes (which are, in fact, inverse slopes measured on the section of the feature which appears linear on a logarithmic plot)

(iii) The height of a peak. This parameter is, in fact, seldom quoted for a depth profile as it is very badly affected by the broadening.

(iv) The energy dependencies of the above parameters.

The often quoted 'depth resolution' Δz, (defined as the depth required for the signal from a step feature to drop from 84% to 16% of the peak value) is based on the approximation/assumption that the response function is Gaussian in form. As the double exponential is often a better approximation (at least for boron in silicon) this parameter is effectively arbitrary, and was not investigated.

6.5.4.2. Method for the simulation of SIMS Data.

It was considered informative to perform these tests across a range of primary ion energies currently in use. Accordingly depth profiles at 2, 4, 6, and 8 keV ion^{-1} were simulated (SIMS depth profiling at energies outside this range is rare). The response functions were measured from MBE sample 10/33 in the EVA2000 instrument at these impact energies, with other analysis conditions shown in table 6.2. The profiles are shown in figure

6.12a where, to allow meaningful comparison across the energy range, we show the data quantified by conventional methods.

Primary Ion Energy / keV ion^{-1}	Primary Ion Current / μA	Frame Time / s	Rastered Area / μm
2	40	50	410
4	100	20	410
6	200	10	402
8	200	10	405

Table 6.2. Analysis conditions for the measurement of Si(B) delta 10/33.

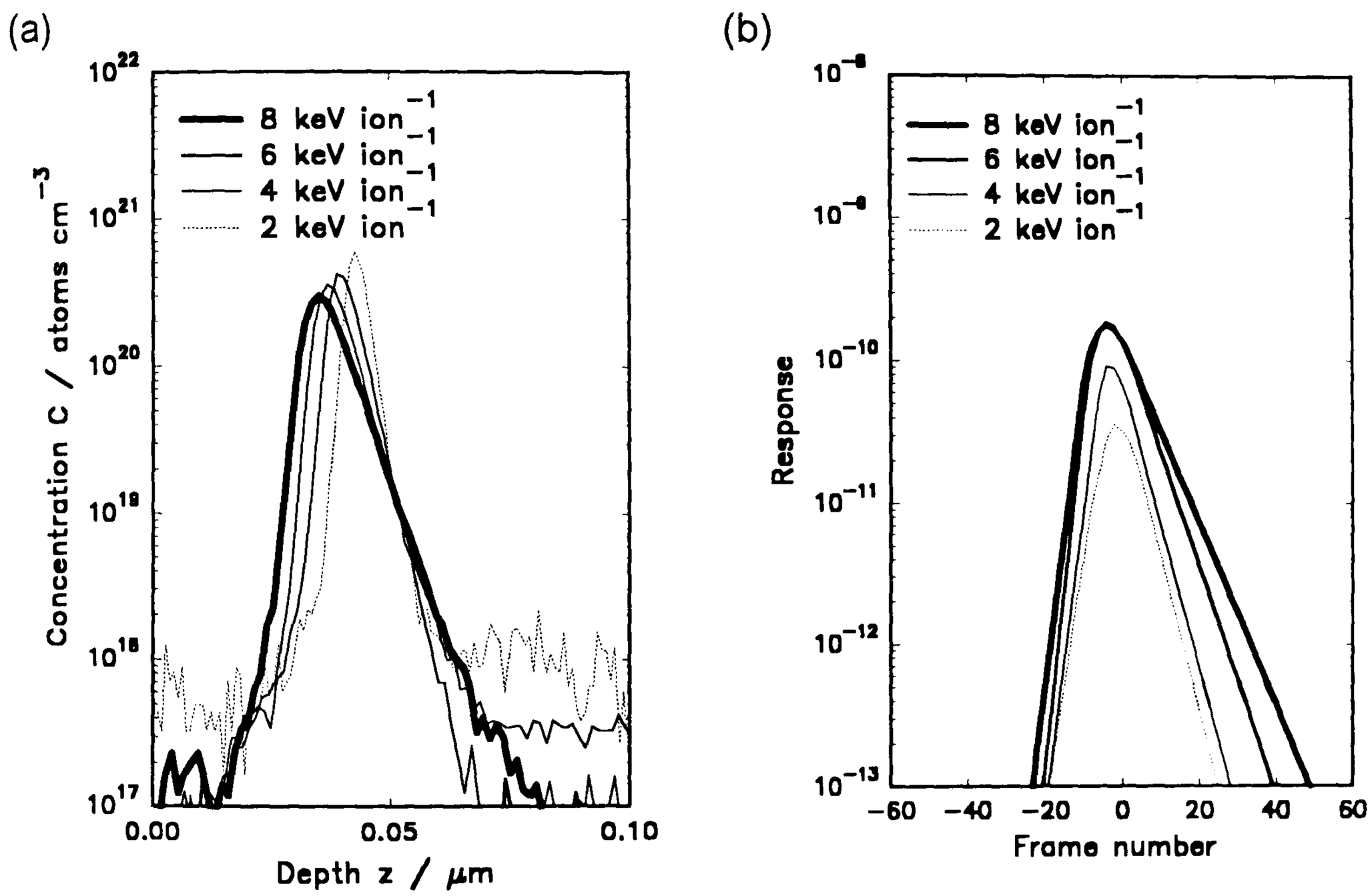


Figure 6.12. Si(B) delta 10/33: (a) calibrated depth profiles, and (b) the response functions.

The response functions were defined as in section 4.2.2.1, taking the differential shift of the peak as -1.3 nm keV^{-1}, as found in the literature (REFS). To determine the effective shift in frames, this shift was scaled using the erosion rate as determined from the delta layer itself (section 4.3). The calculations are shown in table 6.3. The extrapolation method of section 6.3

was used to reduce the effect of the noise on the response. The response functions obtained are shown in figure 6.12b.

Primary Energy E_p / keV	Frame of Peak	Expected shift $\sigma(E_p)$ / nm	Apparent depth z_{app} / nm	Erosion rate $\dot{z}$ / nm frame^{-1}
2	96	-2.6	48.4	0.504
4	56	-5.2	45.8	0.818
6	54	-7.8	43.2	0.800
8	52	-10.4	40.6	0.781

Table 6.3. Erosion rates as calculated from Si(B) delta layer 10/33.

To simulate SIMS data, these response functions were convolved with various assumed concentration distributions described in the next few sections, and Gaussian noise of standard deviation

$$\sigma(f) = 2\sqrt{Y(f) + 5} \qquad (6.12)$$

added. Finally, the profiles were rounded to the nearest integer value.

6.5.4.3. The simulated data set for the FWHM parameter.

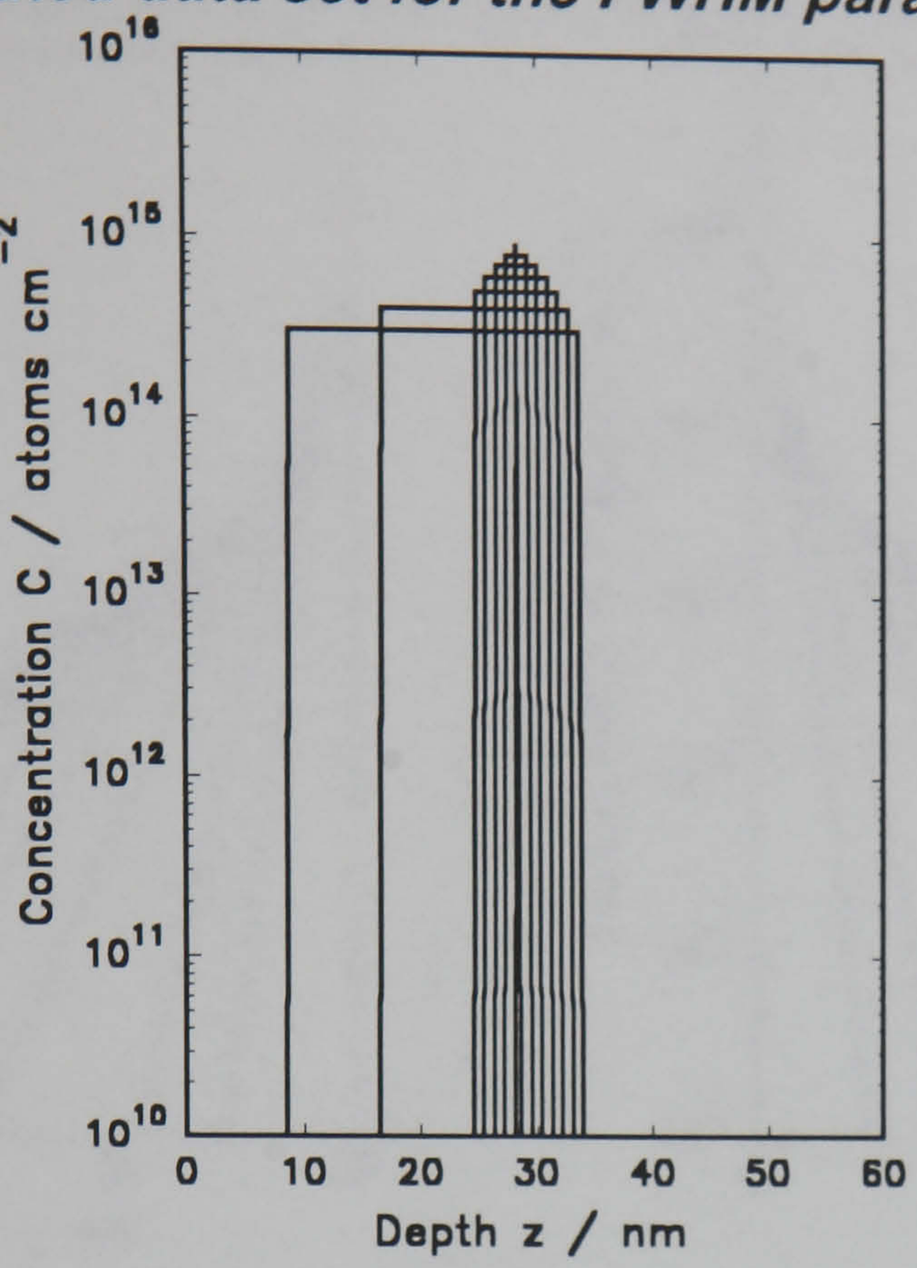


Figure 6.13. Simulated Data Set for full width reconstruction: the concentration profiles, offset vertically for clarity.

The FWHM is one of the more common parameters in the literature. It is related to, and sometimes called, the depth resolution of the technique. Concentration distributions with only this parameter were created: all distributions have a peak concentration of $3x10^{14}$ atoms cm^{-2} over a range from 1 to 30 nm wide. The distributions are shown in figure 6.13, offset vertically for clarity. After convolution with the above response functions and addition of noise, the simulated SIMS depth profiles were depth quantified using the erosion rate, and the FWHM measured using the ProWin quantification software (chapter 8).

6.5.4.4. The simulated data set for the rise and decay inverse slopes.

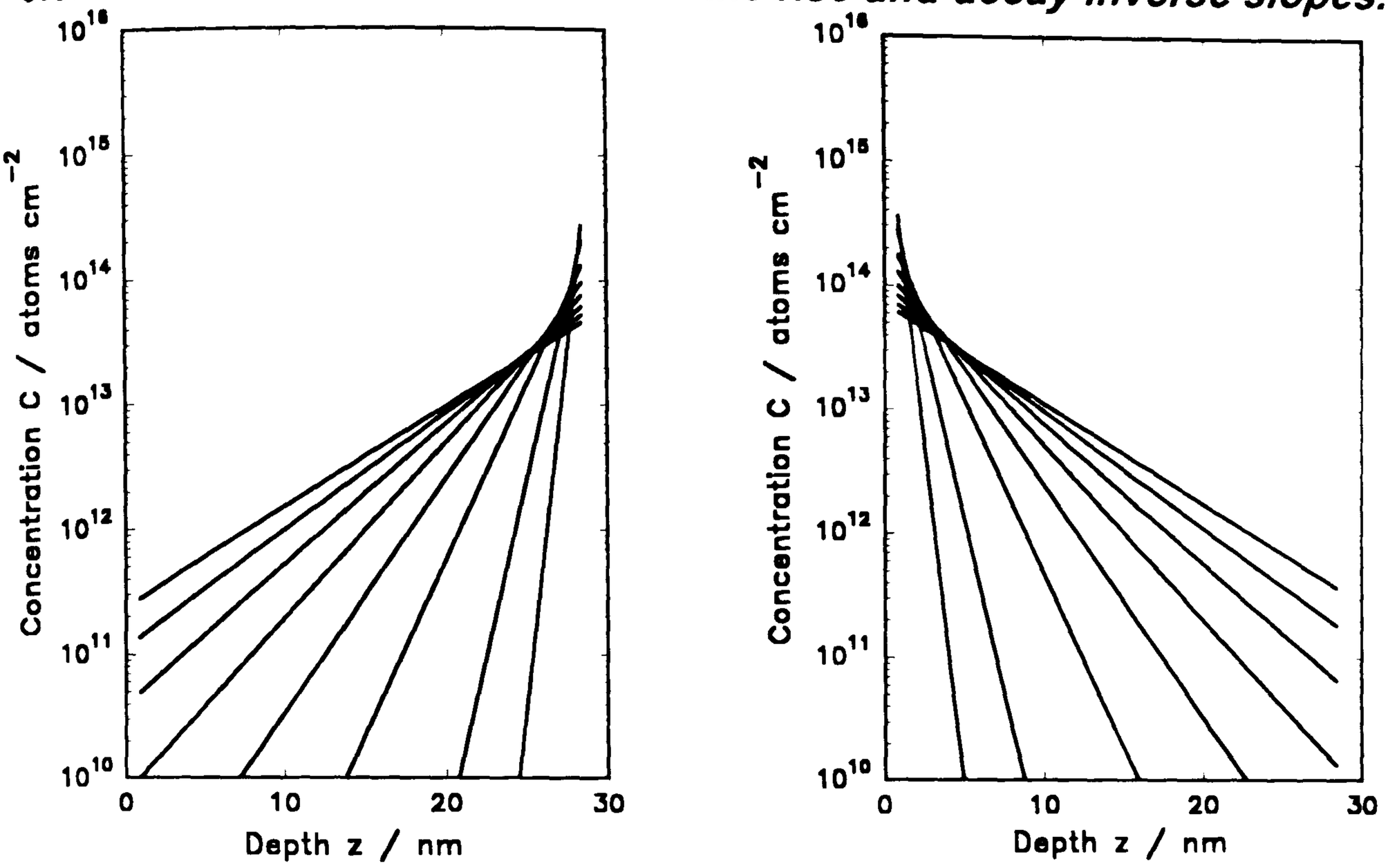


Figure 6.14. Simulated Data Set for slopes reconstruction: the concentration profiles.

To investigate the effect on the slopes of features, concentration distributions consisting solely of a slope (rise or decay) were created, with slopes from 1 to 15 nm decade^{-1}. In order not to bias the results by giving greater statistical precision to steeper slopes, the distributions were normalised to the same integrated concentration. The distributions are shown in figure 6.14. After convolution and added noise, the simulated SIMS data were quantified using the erosion rate, and slopes were measured (using ProWin) from 1% to 50% of the peak value.

6.5.4.5. The simulated data set for the peak concentration.

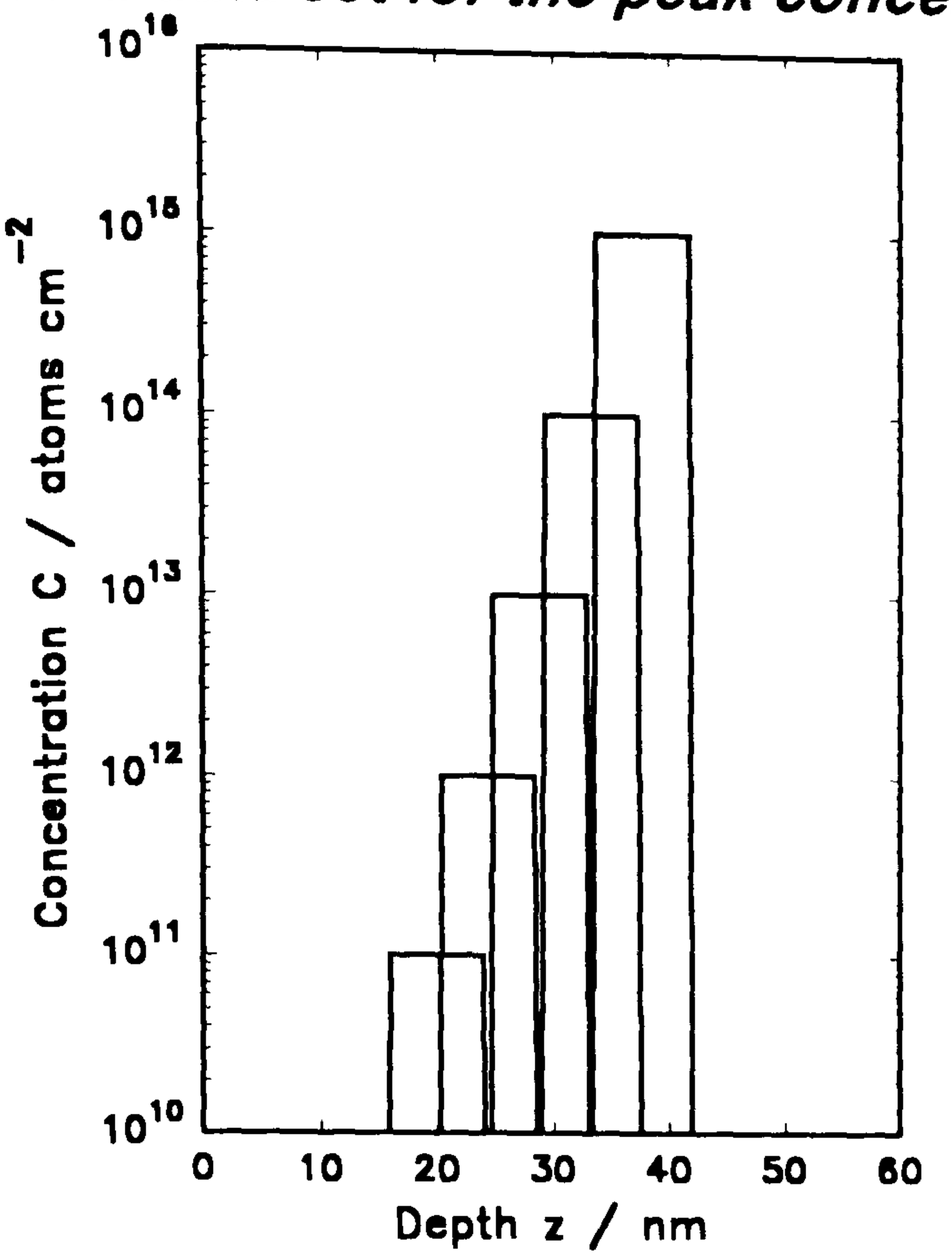


Figure 6.15. Simulated Data Set for peak height reconstruction: the concentration profiles, offset horizontally for clarity.

A parameter which is seldom quoted for SIMS profiles of thin features is the peak concentration, quite possibly because it is one of the parameters worst affected by the blurring. Nonetheless, it is a valuable parameter and we should know when it is reliable, whether we are using conventional quantification or MaxEnt. Again, simple thin layer features were created, approximately 10 nm wide, with peak concentrations from 10^{11} to 10^{15} atoms cm^{-2}. These distributions are shown (offset slightly in the x-direction) in figure 6.15. Again, the parameters were extracted using the ProWin quantification software.

6.5.5. Limiting behaviour of the MaxEnt method.

Having characterised the behaviour of MaxEnt method with regard to various common SIMS parameters, it was decided that it would be informative

to investigate the results of MaxEnt reconstruction of data that was, in some way, limiting.

It was thought that the reconstruction of a genuinely flat distribution would not be sufficiently informative. However, the opposite case was considered worth investigating: the reconstruction of a depth profile of a *genuine delta layer*. The loss of information inherent in the atomic motions, and the presence of experimental noise, imply that there can never be justification for returning an ideal delta as the results of reconstruction. As analysis using higher energy primary ions involves greater information loss, the results of reconstruction were expected to be poorer in these cases. The results (in chapter 7.0) visibly demonstrate the effect of the entropy in the potential function.

For this test it was unnecessary to simulate the depth profiles. The response function measured from a delta layer was used to quantify the profile of the layer. In this way the 'true' profile is an ideal delta function and the response includes all broadening effects *by construction*. For this experiment, the depth profiles of the boron in silicon delta layer 10/33 (described previously) were used.

6.6. Improving the background on the reconstruction.

The results of the MaxEnt reconstruction of test data were disappointing in one respect: the background noise signal is often enhanced as well as the features of interest. Although no deconvolution method can completely avoid these effects, some attempts were made to improve the performance of MaxEnt in this area. These attempts were tested by examining the improvement (or lack of) obtained on reconstruction of the

simulated profiles previously described. Again, the methods are described here, and the results are in chapter 7.0.

6.6.1. Improving the estimate of the standard deviation.

The standard deviation of the noise on each data point, used in the calculation of the misfit statistic, has hitherto been calculated directly from the measured data. This data is itself noisy, resulting in errors in the calculation of σ. Due to the form of the standard deviation equation

$$\sigma \propto \sqrt{m} \tag{6.13}$$

these errors were expected to be proportionately greater at the background level, where each frame is a less reliable estimate of *m*. The effect on the calculated standard deviation is clearly visible in figure 6.16 for the profile of the Si(B) delta layer described in section 6.5.4.2: although smoothly varying over the feature, the standard deviation at the background level varies greatly from frame to frame. The effect is particularly serious in frames which have 'lost' counts due to the noise. The lower value generates a smaller deviation, constraining the reconstruction to a range of count rates which do not contain the 'true' value.

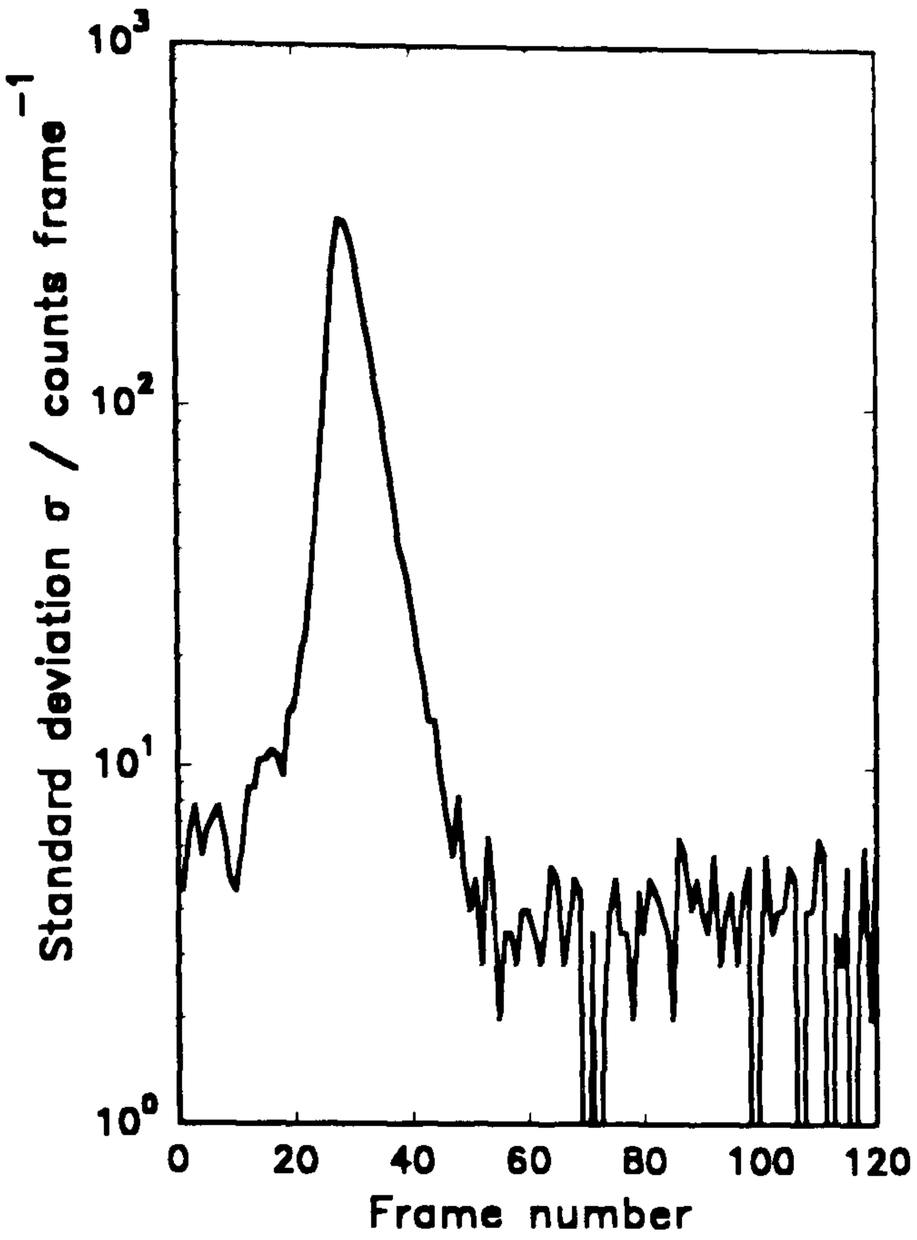


Figure 6.16. The standard deviation of the noise, calculated directly from the measured data.

6.6.1.1. Increasing the estimate of the standard deviation.

The first attempt to correct for this effect was simply to allow slightly greater freedom in the lowest orders by calculating the standard deviation in each frame based on $Y+b$ rather than Y, where b is a small constant, initially ten. The effect of this on the calculated noise deviation is shown in figure 6.17: the calculated standard deviation is significantly smoother in the lowest orders than in figure 6.16, but that on the features is almost completely unaffected.

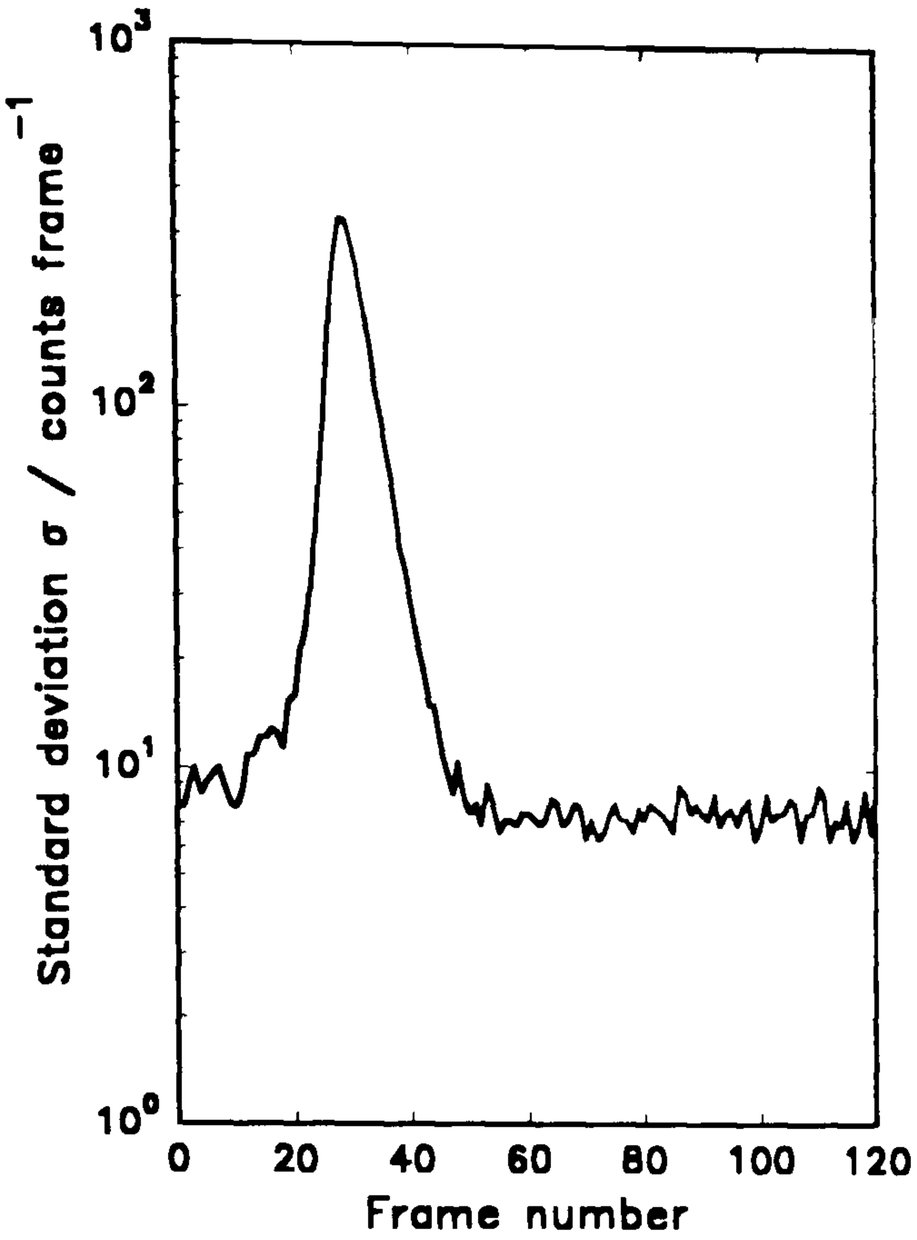


Figure 6.17. The standard deviation, calculated by adding ten counts to each frame, for the same profile as before.

6.6.1.2. The use of smoothing.

A more sophisticated method for improving the estimate of the standard deviation was to use smoothing to obtain a more reliable 'true' value in each frame from which to calculate the standard deviation. It should be stressed that this smoothing was only for the purposes of calculating σ, and did not affect the reconstruction in any other way. The effects of 3- and 7-point smoothing during the calculation are shown in figure 6.18: is clear that, although both give a smoother estimate of the noise deviation at background levels, they also adversely affect (decrease) the calculated deviation on sharp features.

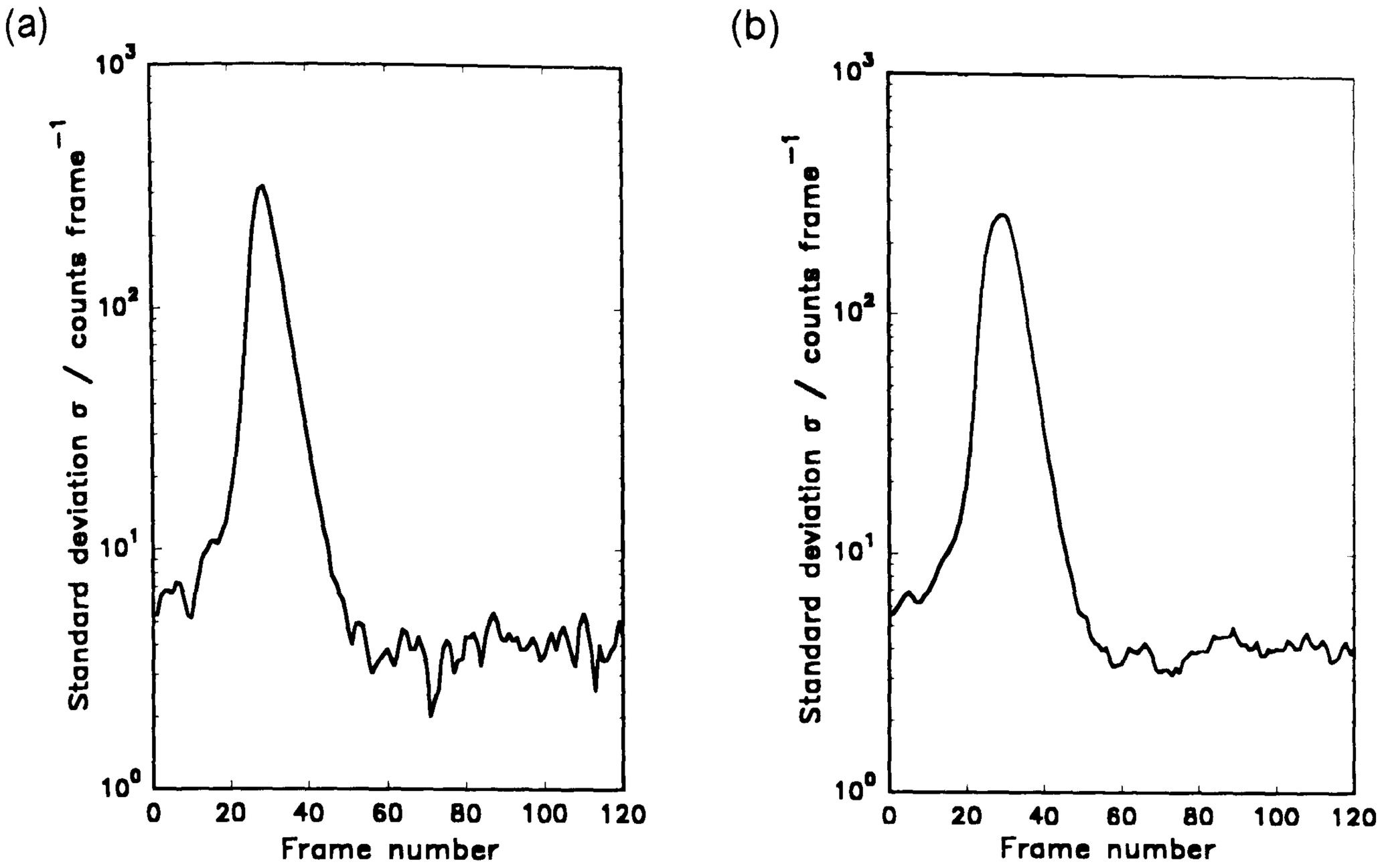


Figure 6.18. The standard deviation calculated using (a) 3-, and (b) 7-point smoothed profiles.

6.6.1.3. Calculating the standard deviation on the simulated profile.

A third alternative considered was not to estimate the standard deviation based upon the measured data, but upon the (noise-free) calculated SIMS profile. As this profile changes during the optimisation, it removes the possibility of pre-computation of $\sigma(f)$, slowing the convergence. It was also considered possible that this change to the objective function could adversely affect the convergence.

All methods described above were tested with a variety of simulated data, and the performance compared to those using the standard estimation method.

6.6.2. A different form of entropy.

The Shannon-Jaynes entropy for a uniform model

$$H(p) = -\int p(z)\log(p(z)), \tag{6.14}$$

where

$$p(z) = \frac{C(z)}{\sum_z C(z)} \tag{6.15}$$

gives particularly little weight to the points in the lowest orders. It was considered that this may allow the search to place low-concentration ordinates almost entirely to match the data (i.e. the background noise) rather than to increase the entropy, resulting in poor noise suppression. To attempt to compensate for this effect, an adjustment to the form of entropy used was sought. A form that appeared particularly suited to SIMS depth profiling was

$$H(q) = -\int q(z)\log(q(z)) \tag{6.16}$$

where

$$q(z) = \frac{\log(C(z))}{\sum \log(C(z))} \tag{6.17}$$

This form matches the bias given by the analyst in plotting the profile on a logarithmic scale. It gives a much greater weight to the lower orders, as can be seen from figure 6.19, where we plot the contribution to the entropy from a single ordinate for both forms of the entropy.

However, there is a theoretical problem with this form of the entropy. The derivation which so elegantly lead to maximum entropy as the most likely distribution is based on the fact that the data are positive and additive (chapter 3). The data are now effectively $\log(C)$, which does not obey this restriction, thus the applicability of MaxEnt is questionable. Nonetheless, it is possible to view the new entropy as an arbitrary regularising parameter: the

method obtained might be justified by its results rather than by its derivation. A number of tests were thus attempted to determine the performance of this form.

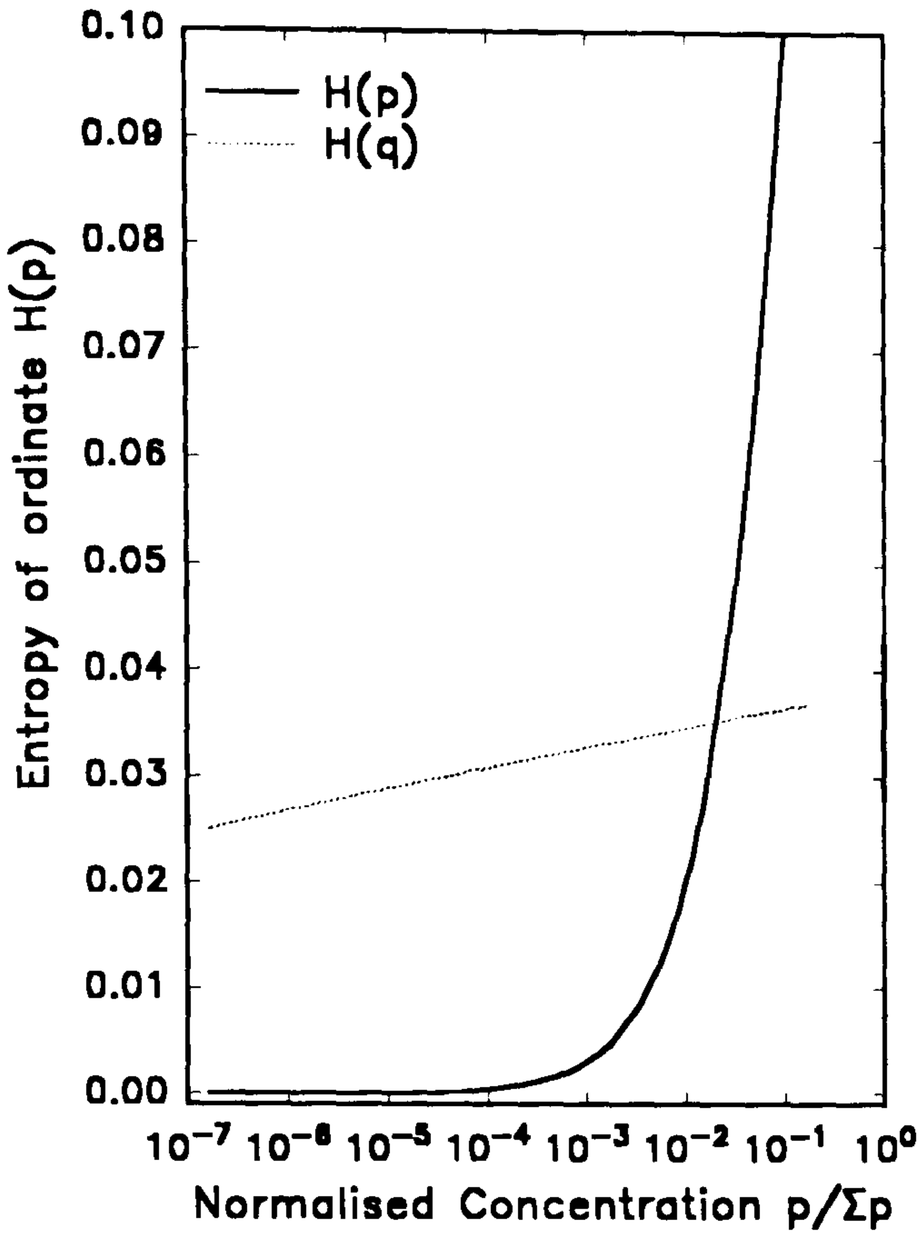


Figure 6.19. The contribution to the entropy of a single ordinate.

6.7. The reconstruction of SIMS data.

It was decided that the MaxEnt method had been suitably tested and characterised that we might progress onto the reconstruction of real SIMS depth profiles. Before any reconstruction was performed it was considered prudent to test the assumption that the measured response would be the same as the response in the unknown sample. As the reconstruction of samples where the true distribution is not known is less informative about the MaxEnt than the use of simulated profiles, it was attempted to limit the samples used to those of particular interest, or those where data was available at a range of primary ion energies where the energy dependence of the reconstructed profiles would give some indication as to the success of the method.

6.7.1. Testing the instrumental stability.

To test that the response measured is still valid during the later profile of the unknown sample, measurements were taken repeatedly from the same delta layer. Firstly, a boron in silicon delta layer, sheet concentration $1.6x10^{14}$ atoms cm^{-2}, at a depth 75 nm, (MBE sample 19/4), was profiled in EVA2000. Operating conditions were: 4 keV $^{16}O_2^+$ at normal incidence and 150 nA, frame time was 10 seconds for a raster of 410 μm, digitally gated to 125 μm. Figure 6.20 shows the results of consecutive profiles. No calibration *and no re-tuning of the instrument was performed* between the profiles: it is a test of the instruments raw stability over one and a half hours (the duration of each profile). The depth resolution parameters, measured directly on the raw data, are shown in table 6.4.

Measured Parameter	First Profile	Second Profile.
Area / counts	$4.1x10^5$	$4.2x10^5$
Peak / counts @ frame	26018 @ 156	26817 @ 156
FWHM / frames	13.5	13.3
Rise slope / frames dec^{-1}	7.9	8.2
Decay slope / frames dec^{-1}	-15.3	-15.9
Erosion rate / nm $frame^{-1}$	0.46	0.46

Table 6.4. Measured parameters from consecutive profiles of delta layer 19/4.

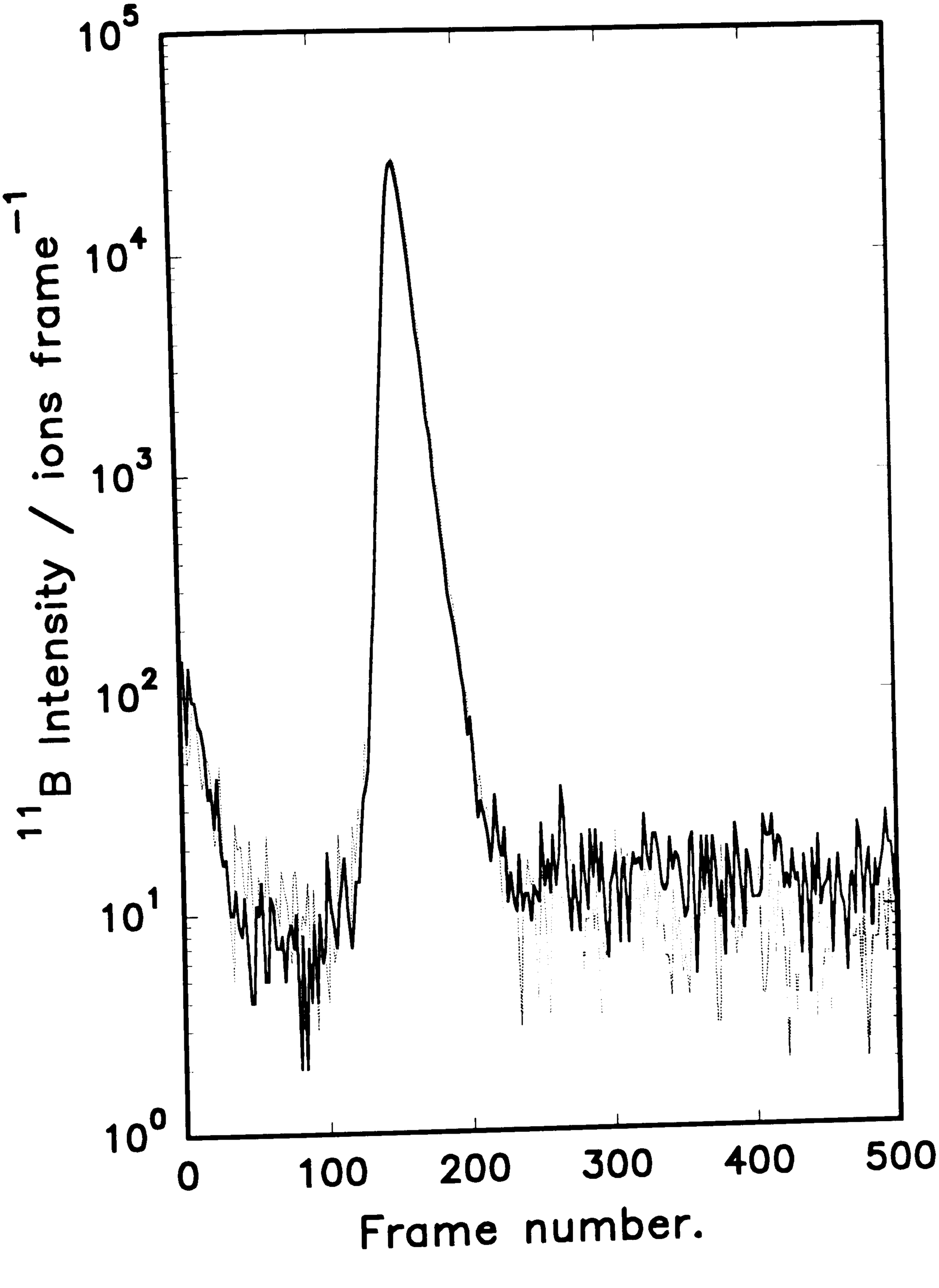


Figure 6.20. Consecutive SIMS depth profiles of a delta layer.

Comparing profiles taken with a greater separation a larger discrepancy was found. The profiles shown in figure 6.21 are of a boron in silicon delta, $3x10^{14}$ atoms cm^{-2}, at 51 nm, (MBE sample 10/33). The profiles were taken in EVA2000, again using 4 keV $^{16}O_2^+$, but the primary ion current dropped slightly during the intervening 15 hours. The measured parameters from these profiles are shown in table 6.5.

Measured Parameter.	**First Profile.**	**Second Profile.**
Area / counts	$7.8x10^5$	$6.9x10^5$
Peak / counts @ frame	53151 @ 96	42862 @ 102
FWHM / frames	12.5	13.7
Rise slope / frames dec^{-1}	6.1	6.8
Decay slope / frames dec^{-1}	-14.9	-16.0
Erosion rate / nm $frame^{-1}$	0.46	0.44

Table 6.5. Measured parameters from profiles of delta layer 10/33 separated by fifteen hours.

It was concluded from this that a measured response function is valid only for the next sample profiled *unless specific steps are taken* to ensure that the instrumental conditions are identical. Generally it will be difficult to ensure that a re-tune of the instrument gives identical conditions, thus it is preferable to limit the 'useful life' of a response function. This situation, although a little disappointing, was not unexpected as it is currently the case for conventional quantification methods (section 2.2).

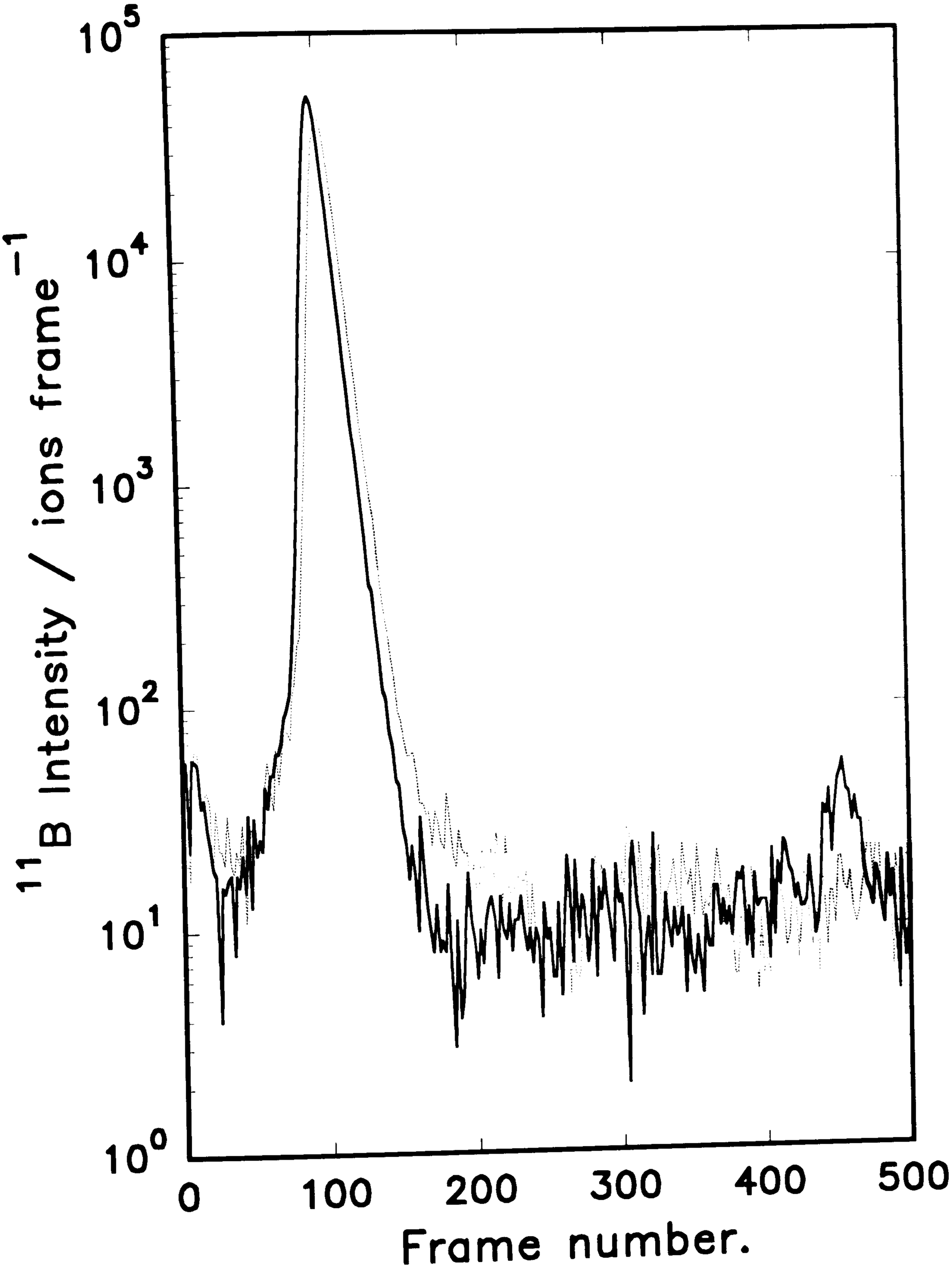


Figure 6.21. SIMS depth profiles of a delta layer separated by fifteen hours.

6.7.2. Transferring response functions.

It was further attempted to check that the response determined from different samples, profiled under identical conditions, is the same. Response functions were extracted from the profiles of the two delta layers described in the last section. These samples were grown and analysed separately. The two functions are shown overlaid in figure 6.22: it can be seen that the agreement is fair.

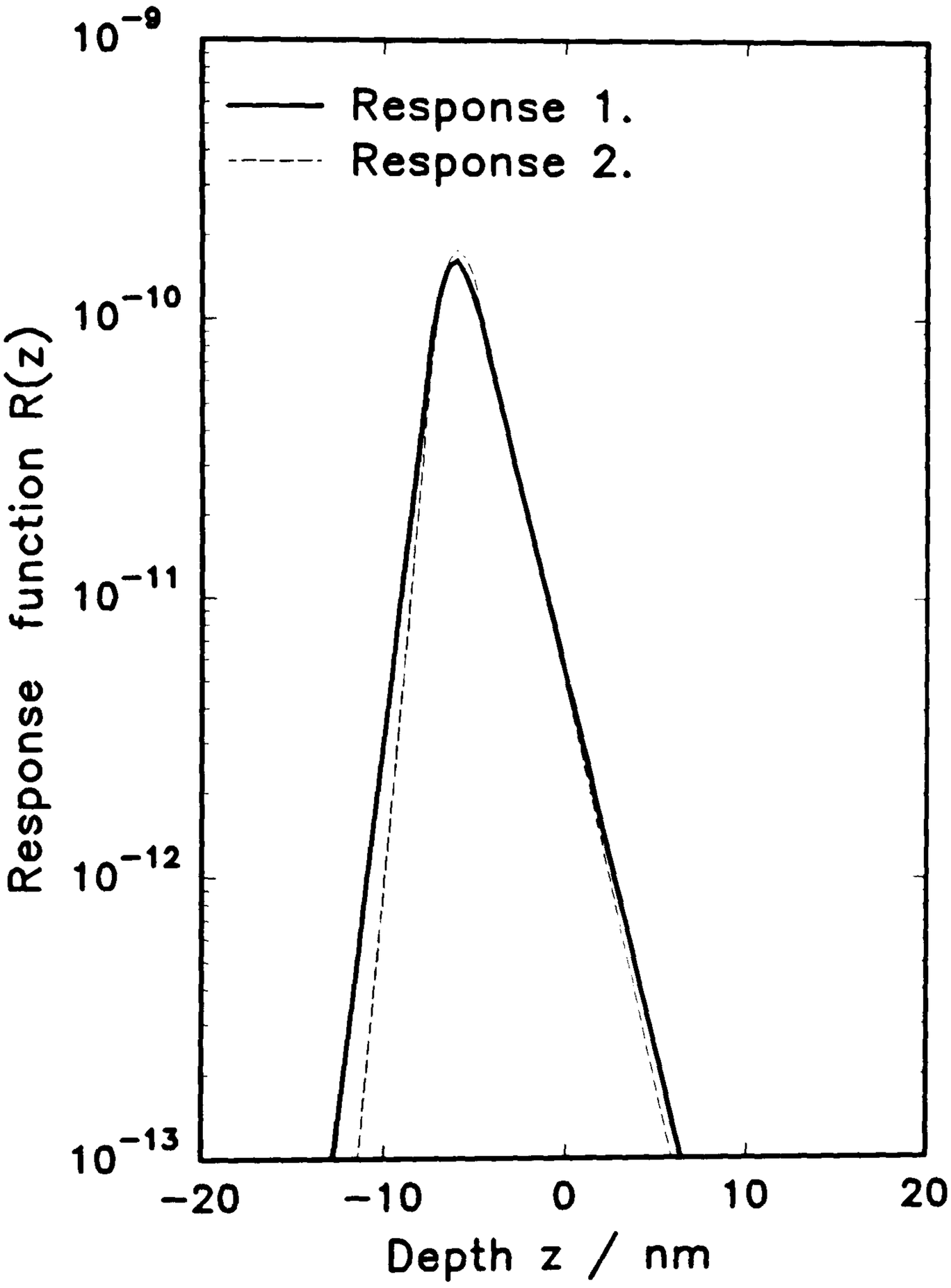


Figure 6.22. Response functions from two different boron delta layers.

The results of the two previous experiments indicated that the reconstruction of SIMS depth profiles was valid, provided some care was

taken during the analysis. We thus attempted the reconstruction of some unknown samples. Such experiments may be significantly less instructive, thus it was attempted to select samples for which some testing of the MaxEnt method could be performed. These tests can best be performed by comparing the results of reconstruction at different bombardment energies.

6.7.3. SIMS Data Set 1: Si Segregation in GaAs.

A process of some interest to growers is the segregation of impurities at the surface of the sample during growth. This detrimental effect exists for some combinations of impurity/sample/growth conditions. If the samples are analysed using SIMS, it may be difficult to infer the amount of segregation during growth due to the atomic motions during analysis.

To investigate this effect, depth profiles were obtained from Prof. M. Maier at the Fraunhofer-Institut fur Angewandte Festkorperphysik, Germany. Two samples had been grown using solid-source MBE to investigate the mechanisms involved in the growth of GaAs:Si heteroepitaxy. After a buffer layer of GaAs, Si was deposited at 450°C, then capped with GaAs at 580°C. For the first sample, the thickness of deposited Si was 0.1 nm, for the second it was 0.7 nm. The amount of Si in the samples had been analysed by High Resolution X-ray Diffractometry, and the lattice-strain patterns determined using High Resolution Electron Microscopy (Brandt et.al, 1993). For the extraction of the response function of the instrument, a Si delta-doped layer in GaAs was used, 7.7×10^{13} atoms cm^{-2} at a depth of 100 nm.

The 0.7 nm sample was profiled at 6 and 15 keV ion^{-1}, the 0.1 nm sample at 6 keV ion^{-1}. Experimental conditions for both energies were: Cs^+ ions, 50 nA at 45°, frame time was 1 second for a 500 μm square crater,

digitally gated to the central 30%. Secondary $AsSi^-$ ions were detected. Raw profiles are shown in figure 6.23, clearly showing the segregation of Si during growth. In figure 6.23a the difference between the 6 and 15 keV samples clearly shows the differential shift and increased broadening at higher energy. Analysis conditions for the response function were identical, raw plots are shown in figure 6.24.

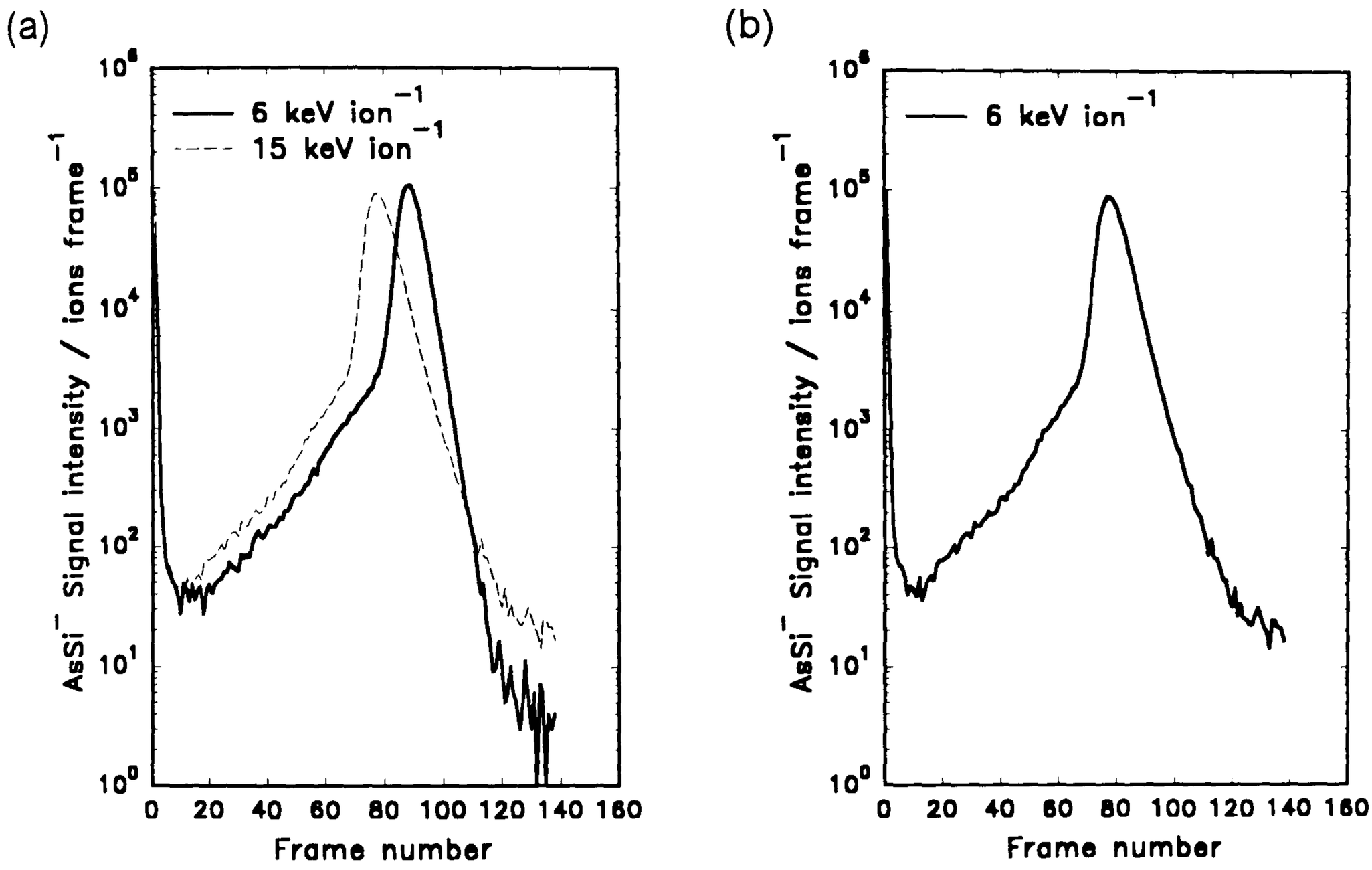


Figure 6.23. SIMS Data Set 1: (a) and (b) the GaAs(Si) thin layers exhibiting segregation during growth.

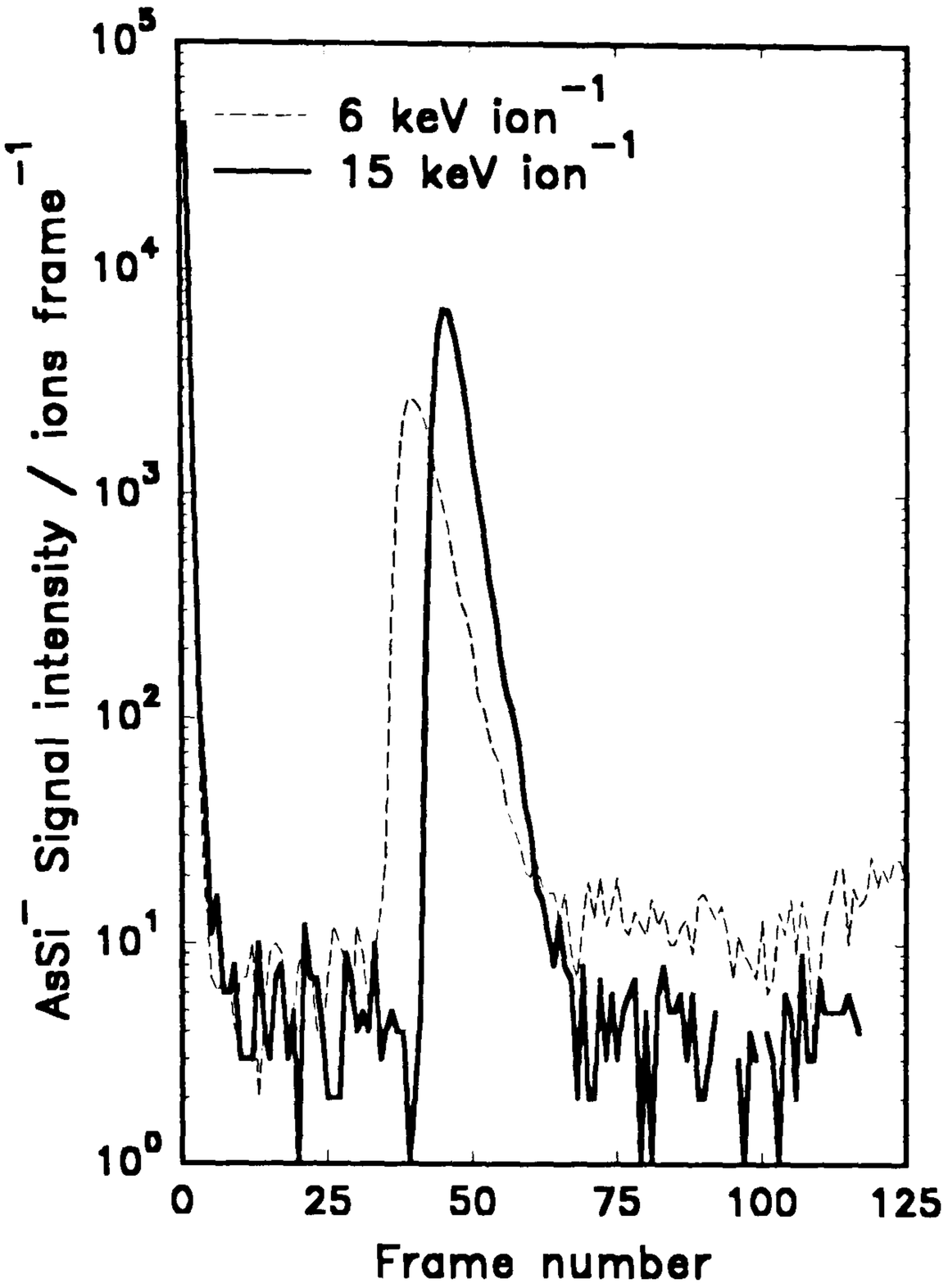


Figure 6.24. SIMS Data Set 1: the GaAs(Si) delta layer used to measure the response, at 6 and 15 keV ion^{-1}.

The large surface Si signal visible on the depth profiles indicated that the pre-equilibrium conditions for these samples are very different from equilibrium. Accordingly, the first 10 frames were neglected as pre-equilibrium, the next 128 data points used. The response function was obtained as in section 6.3, extrapolation using a line of best fit on the logarithmic plot to reduce noise. As knowledge of the differential shift for these samples was not available, no shift was assumed other than that due to the asymmetric nature of the response. Accordingly the response function was placed about its centroid. The results of reconstruction may be found in section 7.4.1.

6.7.4. SIMS Data Set 2: Si(B) triple-delta structure.

The second genuine SIMS data set used was one for which a significant improvement in quantification accuracy was expected. The sample was nominally a boron in silicon triple delta structure, sheet concentrations of $2x10^{13}$ atoms cm^{-2} at 50, 80, and 130 nm. It was grown using MBE in the modified VG Semicon V80 at Warwick (MBE sample 10/34). It has been previously profiled using SIMS by Mr. R.D. Barlow at Warwick, a quantified profile is shown in figure 6.25. To perform the reconstruction the response was measured from the boron in silicon delta 10/33, grown immediately *prior* to this sample.

Primary Energy / keV ion^{-1}	Current / nA	Frame time / s	Raster size / μm
9	500	5	404
6	250	5	402
4	150	10	410
3	120	10	410

Table 6.6. Analysis conditions for measurement of the Si(B) triple-delta structure.

Both samples were profiled in the EVA2000 instrument by Mr. R.D. Barlow over a range of bombardment conditions as shown in table 6.6. As profiles at a range of primary ion energies are available, this experiment did allow a comparison of the reconstruction result with SIMS data taken at a lower energy.

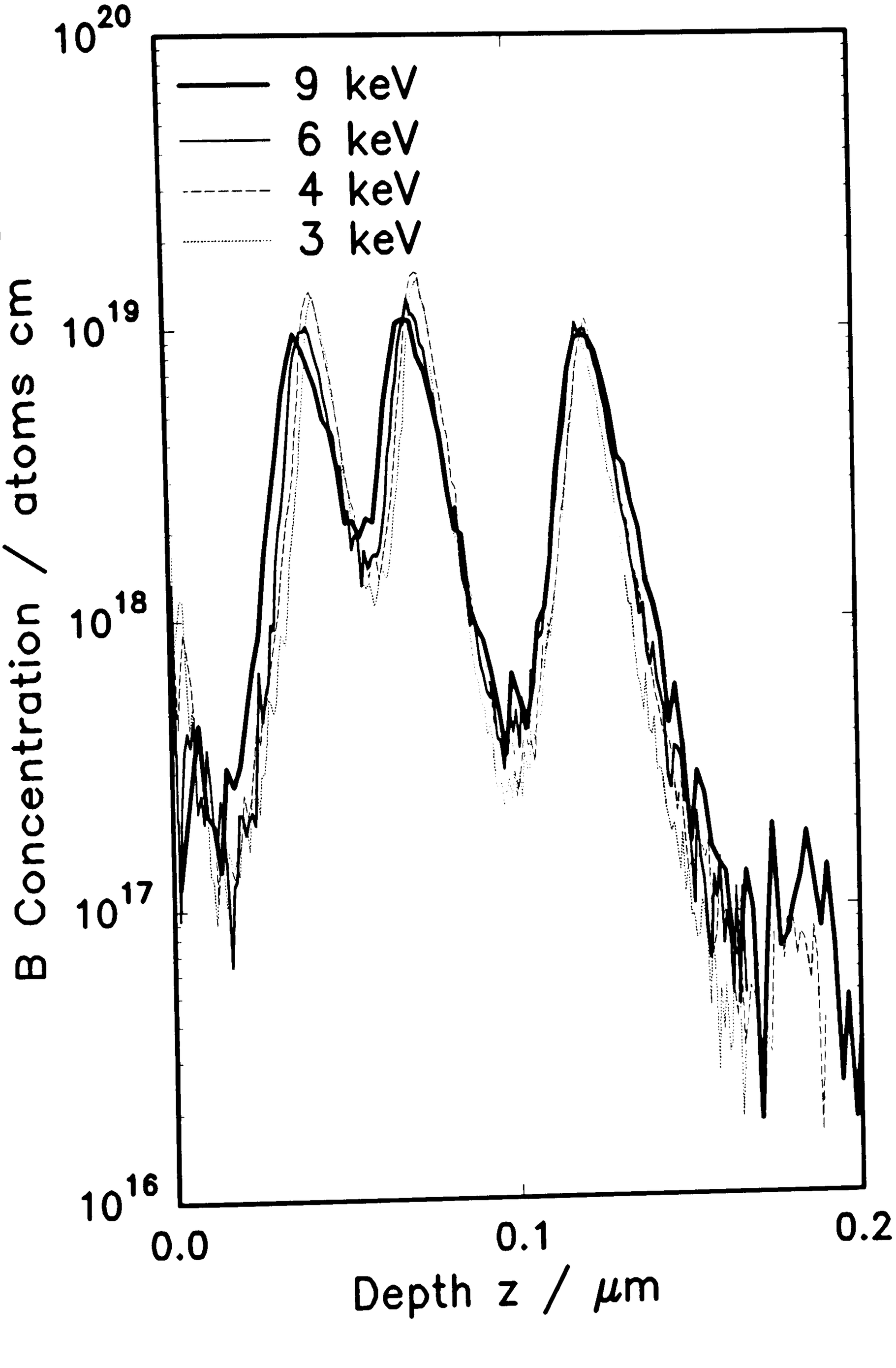


Figure 6.25. SIMS Data Set 2: Raw SIMS depth profiles of the Si(B) triple delta structure.

Analysis conditions for measurement of the response function were the same as for the profiling of the triple delta sample. Extraction of the response function from the profiles was performed as in section 6.3.2. As the layers are boron in silicon, the response was the placed to match the equivalent (in frames) of the expected differential shift of -1.3 nm keV^{-1}. For these purposes the erosion rate was deduced from the delta layer itself, using equation 4.21. The calculation is shown in table 6.7.

E_p / keV ion^{-1}	Shift σ / nm	Expected depth z_p / nm	Occurred at f_c / frames	Erosion rate $\dot{z}$ / nm frame^{-1}
9	11.7	39.3	42	0.94
6	7.8	43.2	94	0.46
4	5.2	45.8	92	0.50
3	3.9	47.1	118	0.40

Table 6.7. Erosion rates for the Si(B) delta, determined from knowledge of the differential shift.

As a test of the method for calculating the erosion rates in this way, these erosion rates were compared with those as determined by crater depth measurement (Sloane Dektak 30/30 Auto 2), shown in table 6.8.

E_p / keV ion^{-1}	Measured Depth z / nm	No. Frames f / frames	Erosion rate $\dot{z}$ / nm frame^{-1}
9	187.3	200	0.94
6	152.65	330	0.46
4	135.3	263	0.51
3	99.8	249	0.40

Table 6.8. Erosion rates for the Si(B) delta, determined by crater depth measurement.

The response functions are shown in figure 6.26.

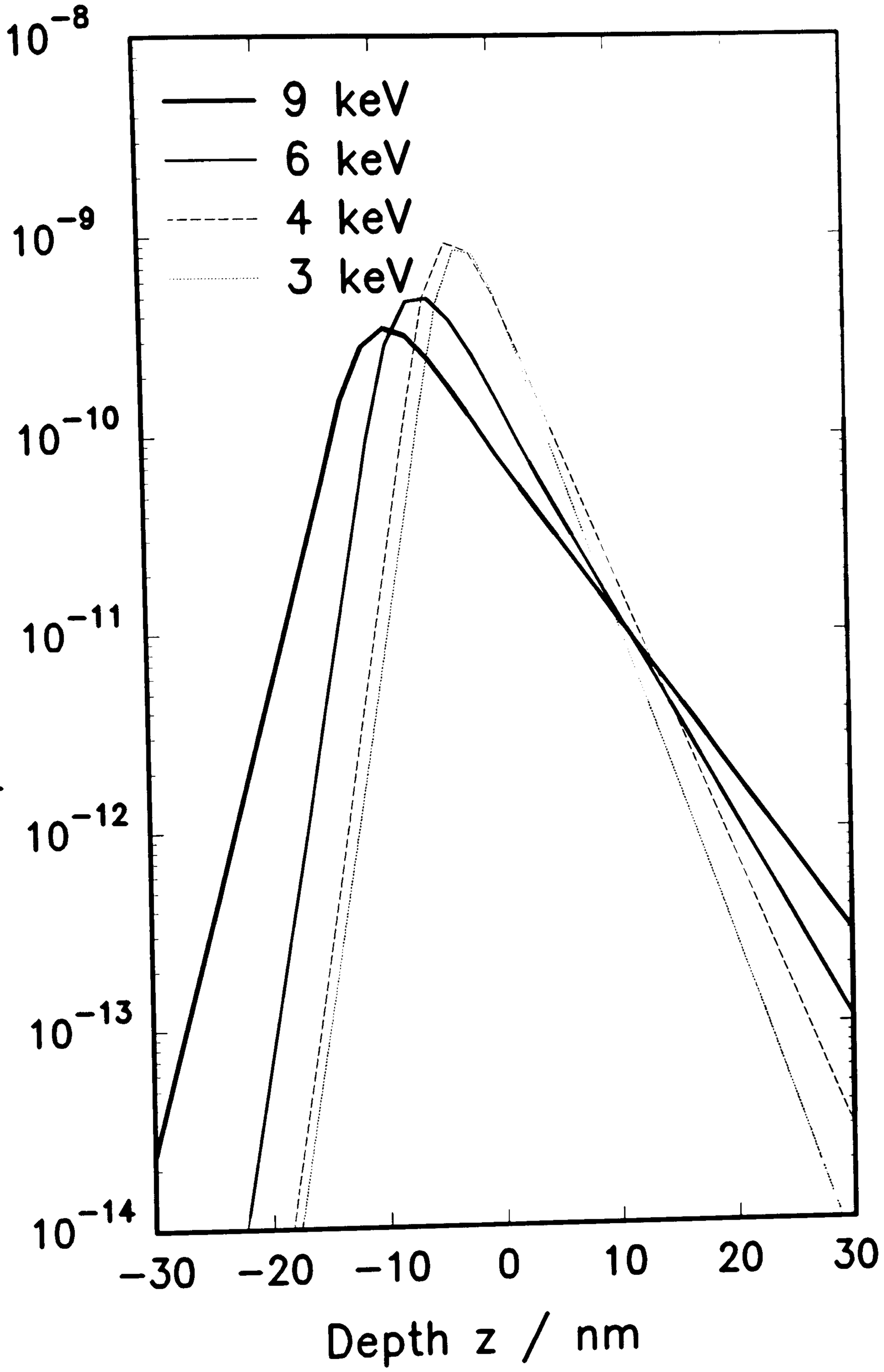


Figure 6.26. SIMS Data Set 2: Response functions as measured from sample 10/33.

7.0. Results and discussion.

In this chapter we shall present and discuss the results of the experiments described in chapter 6. The order of the results follows that of the experimental section: the characterisation of the noise on the depth profile; the appraisal of the algorithms used; the results of tests using simulated depth profile data; and finally the results of the reconstruction of genuine SIMS depth profiles.

7.1. The noise on a depth profile.

7.1.1. The preliminary investigation.

7.1.1.1. The matrix channel.

The preliminary investigation used the silicon channel from available SIMS depth profiles of an implanted boron in silicon standard under a variety of bombardment conditions (section 6.2.3.1). Plotting histograms of the occurrence of each intensity value using method 6.2.2.1 (assuming the signal is constant) demonstrated that, for many profiles, the distributions were very well approximated by a Gaussian. An example is shown in figure 7.1.

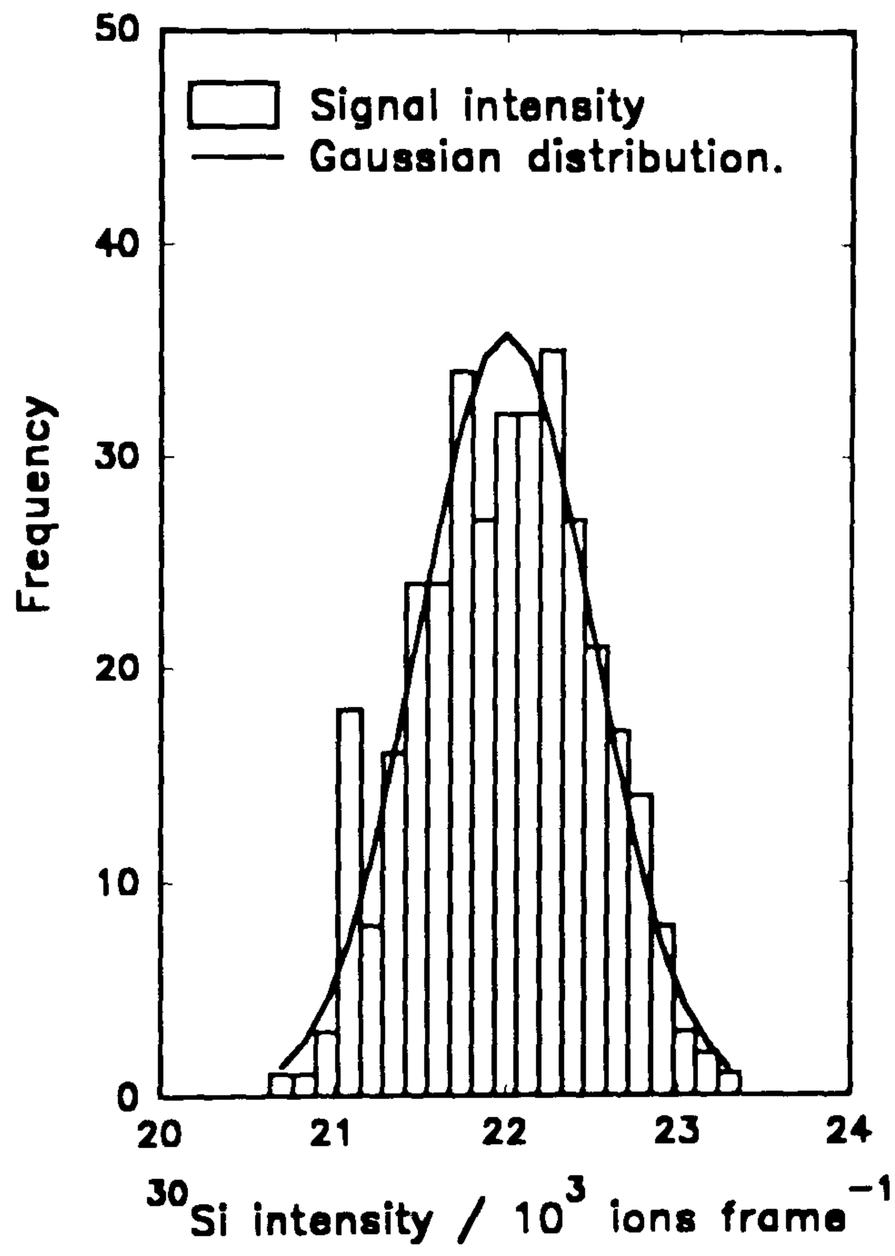


Figure 7.1. Noise distributions: intensity values on a silicon matrix channel.

For some profiles, however, the distributions were found to be particularly asymmetric, or contained multiple peaks. In these cases it was

generally found that the matrix signal was not stable: either a number of stable regions existed, or a slow change occurred continuously during the profile. In figure 7.2a we show the distribution from a profile where the primary ion current dropped during the analysis. In these cases the distribution was investigated using the method of section 6.2.2.2 (fitting a straight line to the data, and calculating the residuals). Generally a Gaussian distribution was then obtained, as shown in figure 7.2b.

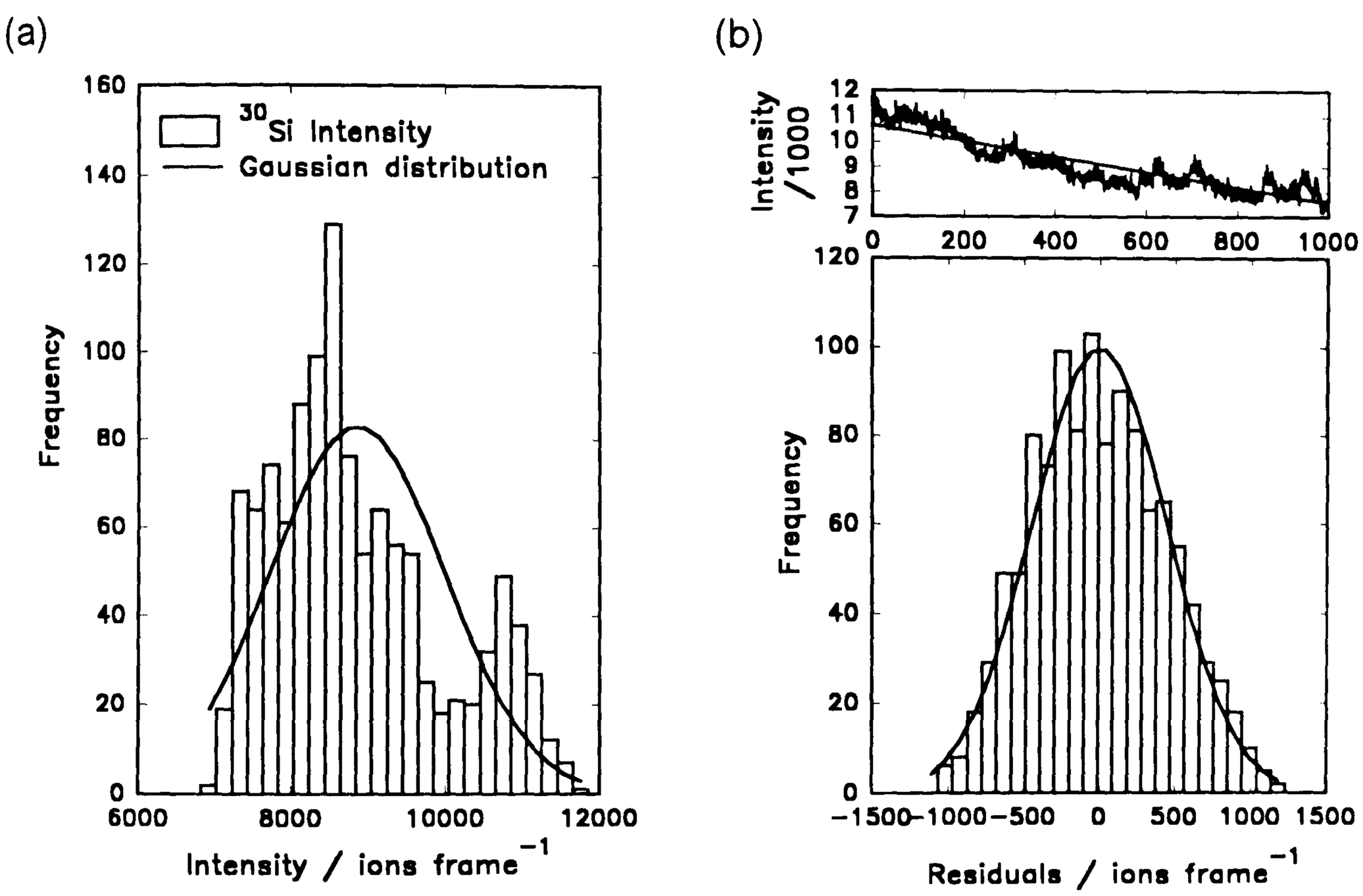


Figure 7.2. Noise distributions: (a) for a slow drop in signal, and (b) the residuals after a linear fit.

In the few cases where this method failed to yield a Gaussian distribution, it was found that the use of method 6.2.2.3 did. However, the use of methods 6.2.2.2 and 6.2.2.3 was discontinued for two reasons:

(i) The fitting of a form to the data may impose some bias toward a symmetric distribution of residuals, leaving the results difficult to interpret, and

(ii) It was decided that the determination of the relationship between the standard deviation and the mean signal intensity was of more importance. The use of either method makes this investigation impossible.

Using only those results from profiles where the signal intensity was stable, the relationship between the mean value and the standard deviation was investigated. The predicted relation is shown in figure 7.3; the straggle of the data is large thus even a linear regression is hard to justify. Nonetheless, it is clear that extrapolation of this data down toward the lower signal intensities likely for dilute impurities is meaningless: the predicted standard deviation for small signals of 200 is far too large.

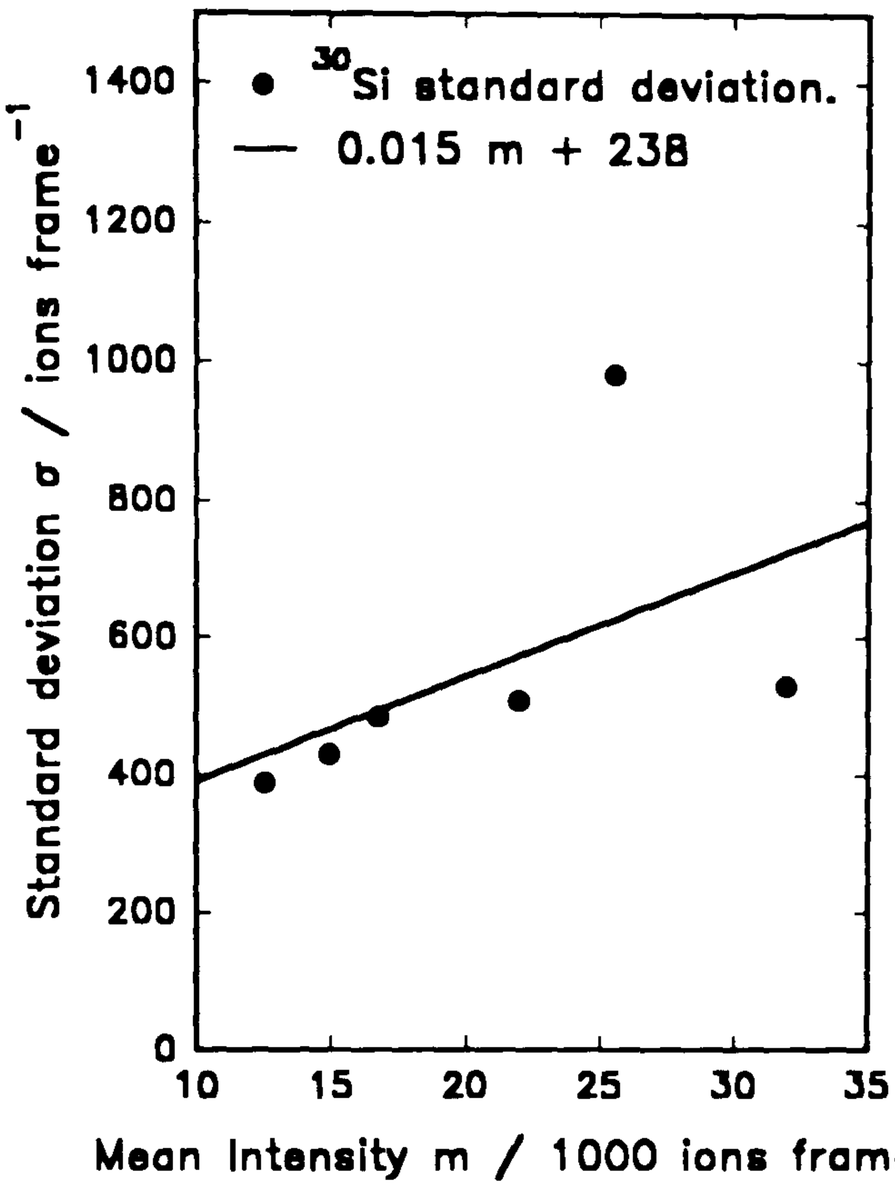


Figure 7.3. The relation predicted by the preliminary study: Si matrix.

7.1.1.2. The background signal.

Typical results of the preliminary investigation of the background noise using thin layers are shown in figures 7.4a and 7.4b. This investigation indicated that a Poisson distribution would be appropriate for the background signals of either boron in silicon or silicon in gallium arsenide, but demonstrated that the Gaussian distribution is not inaccurate for mean signals above 10 counts frame^{-1}.

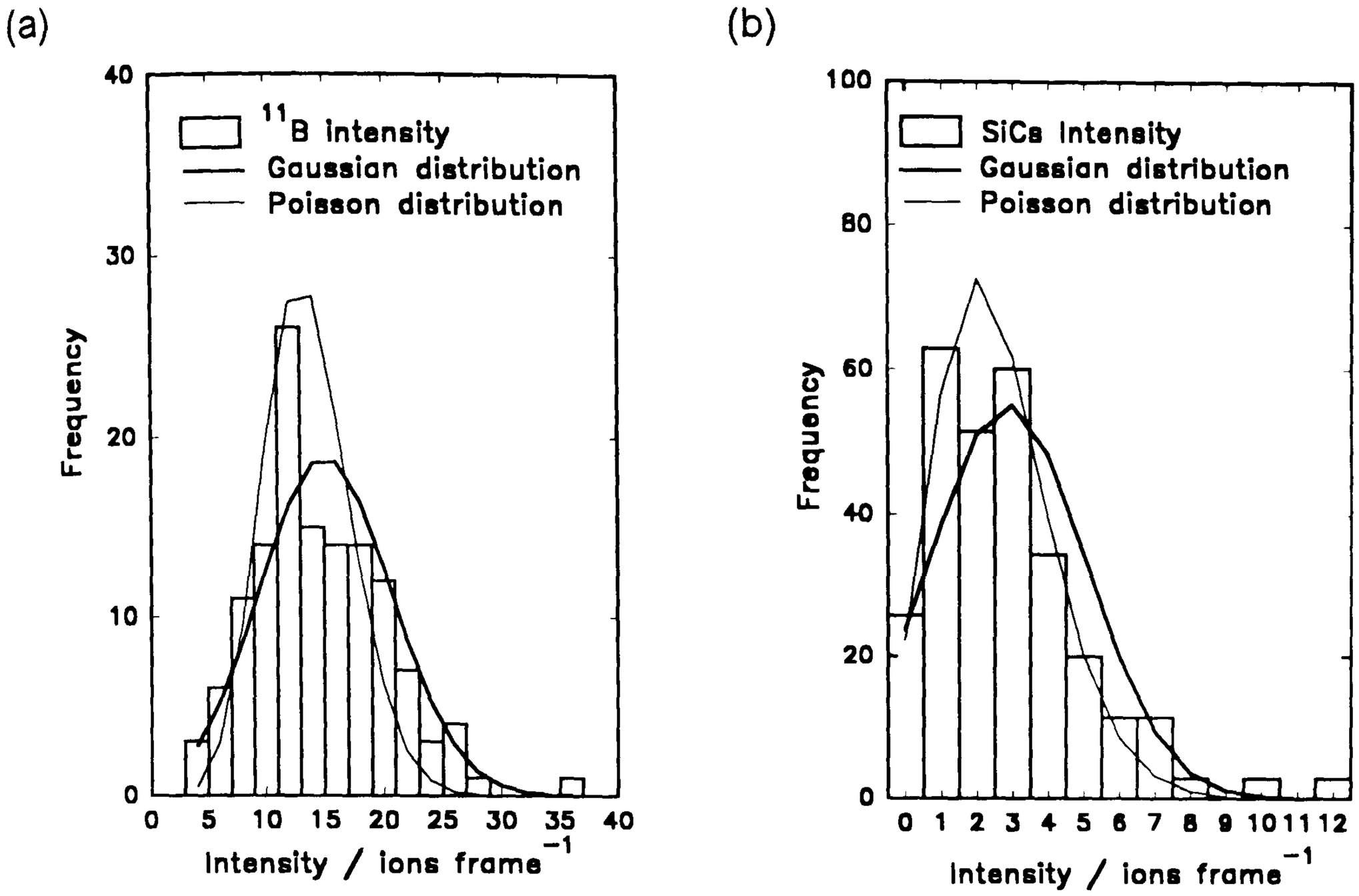


Figure 7.4. Noise distributions: background signals of (a) boron in silicon, and (b) silicon in gallium arsenide.

Plotting the standard deviation of the Gaussian distributions as a function of the mean value the graph shown in figure 7.5 was obtained. The data are reasonably well matched by a linear fit. However, extrapolation into the higher signal region predicted a standard deviation of approximately 5000 at a mean count rate of 20000; in disagreement with the results of the matrix channel.

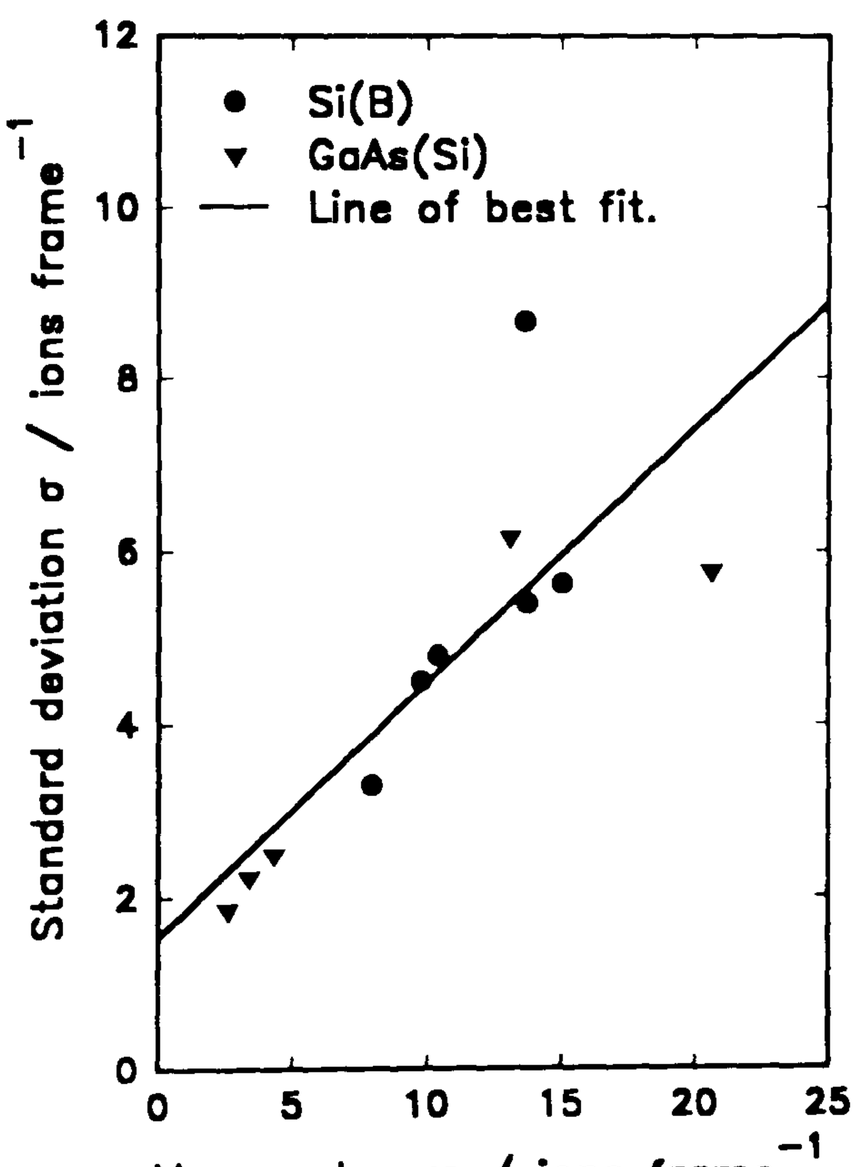


Figure 7.5. The relation predicted by the preliminary study: background noise.

7.1.2. Investigation using a staircase sample.

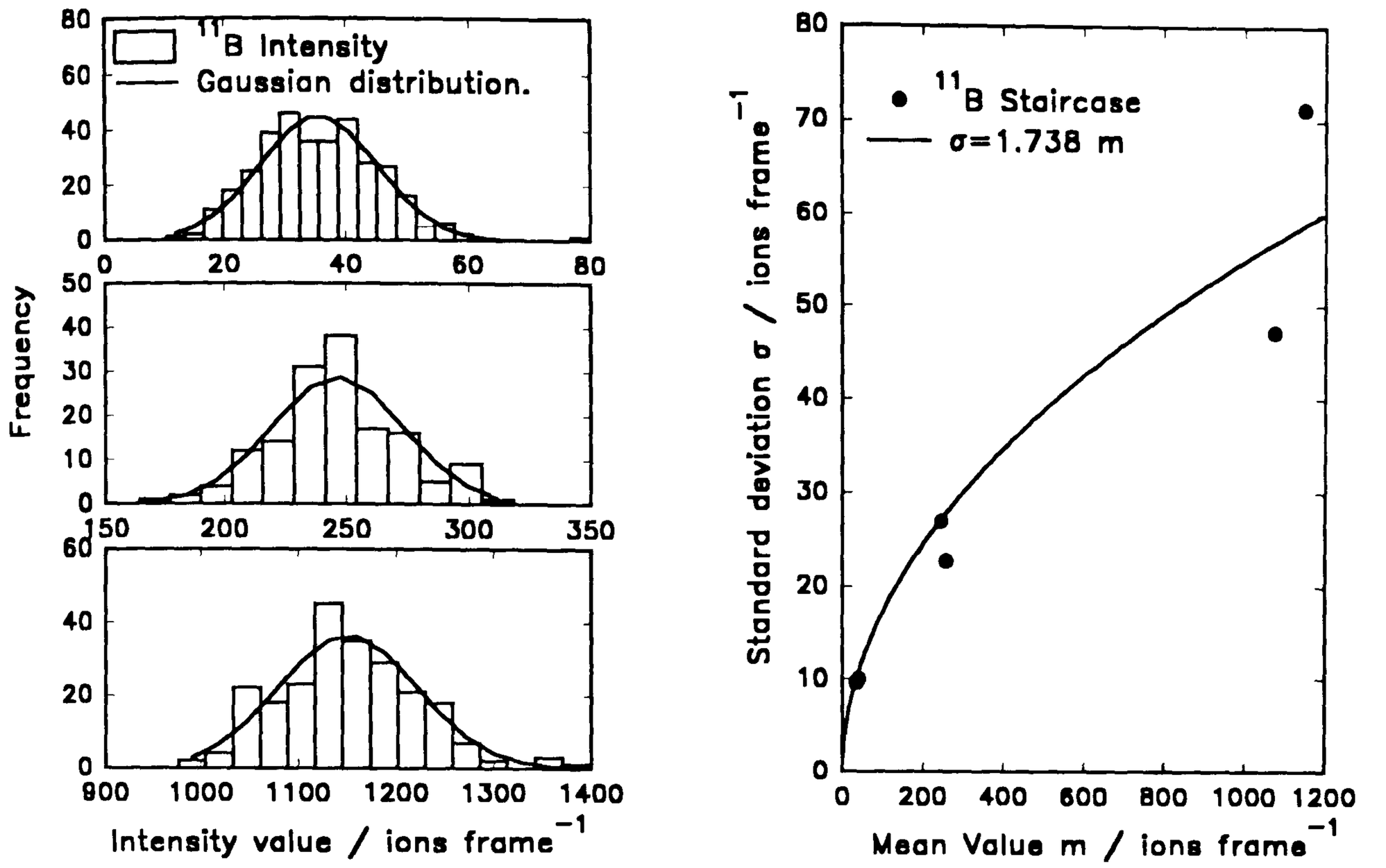


Figure 7.6. (a) The noise distributions on the steps of the Si(B) staircase, and (b) the relation predicted.

Investigating the distribution of the intensity values on the steps of the boron in silicon staircase sample 17/3, it was found that the distributions were all well matched by a Gaussian with suitable mean and standard deviation. Figure 7.6a shows one of the two sets obtained from this sample. Plotting standard deviation versus the mean value of these Gaussian distributions, the graph shown in figure 7.6b was obtained. These values span the range between the two preliminary studies, and the lowest order polynomial to fit the data reasonably was the square-root dependence shown. This dependence follows that expected from the Poisson distribution,

$$\sigma \propto \sqrt{m} \qquad (7.1)$$

except for the constant factor of 1.738. A likely explanation was recently discovered, in the form of a poor earth connection on one of the racks. This

earth connection had probably gone unnoticed for a number of years, and after the repair the stability of the reference signals visibly improved (Cooke, private communication). As yet no experiments have been performed to quantify the change in noise.

7.1.3. Investigation of the energy dependence.

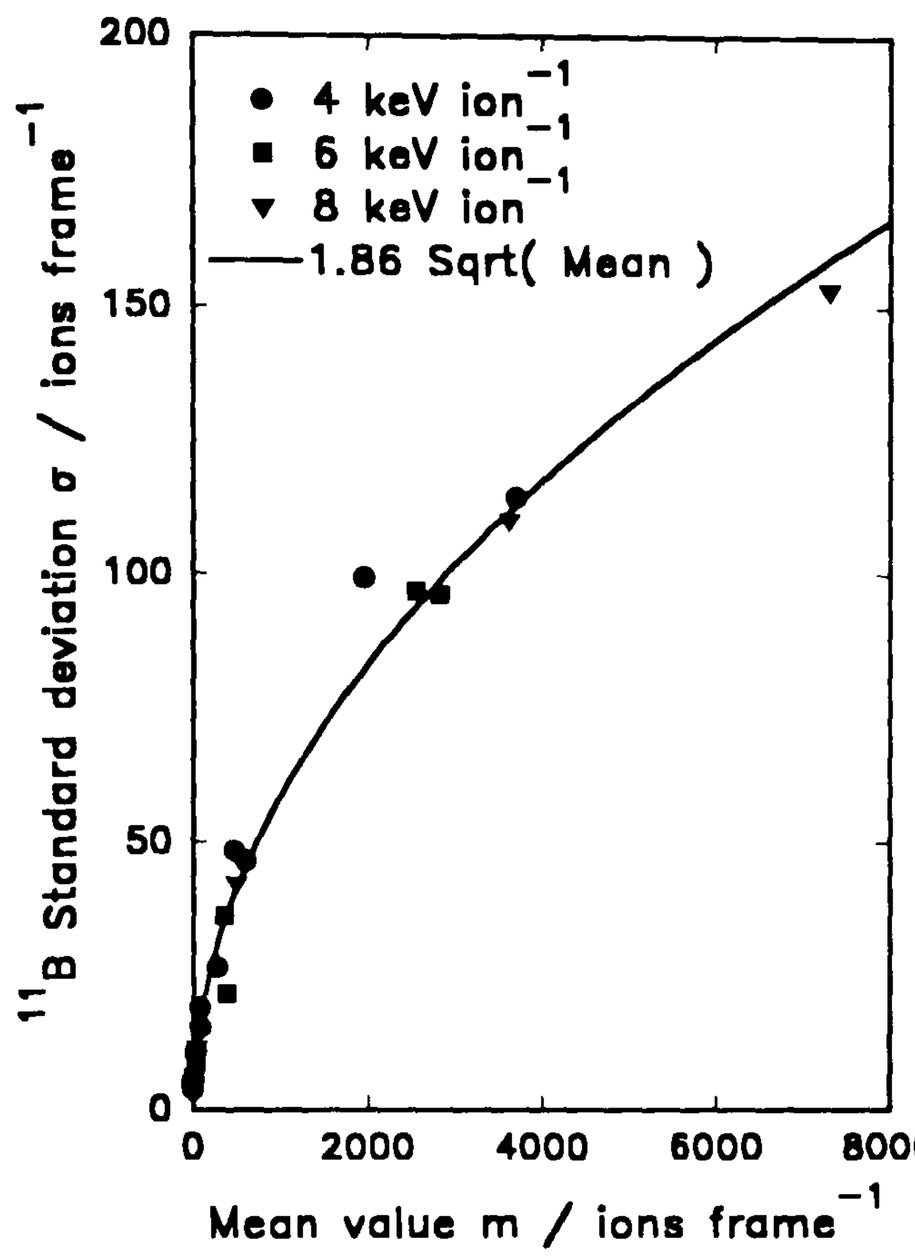


Figure 7.7. The relation for the Si(B) staircase sample 33/02 profiled at a range of primary ion energies.

On repeating the noise characterisation at a range of primary ion energies, the results shown in figure 7.7 were found. The Si(B) staircase used was MBE sample 33/02, profiled at 4, 6, and 8 keV ion^{-1} (section 6.5.2). The results covered a far wider range of mean values than previous studies, and agreed well with the previous staircase study (figure 7.6b). There appeared to be no dependence on the primary ion energy.

7.1.4. Investigation of the sample dependence.

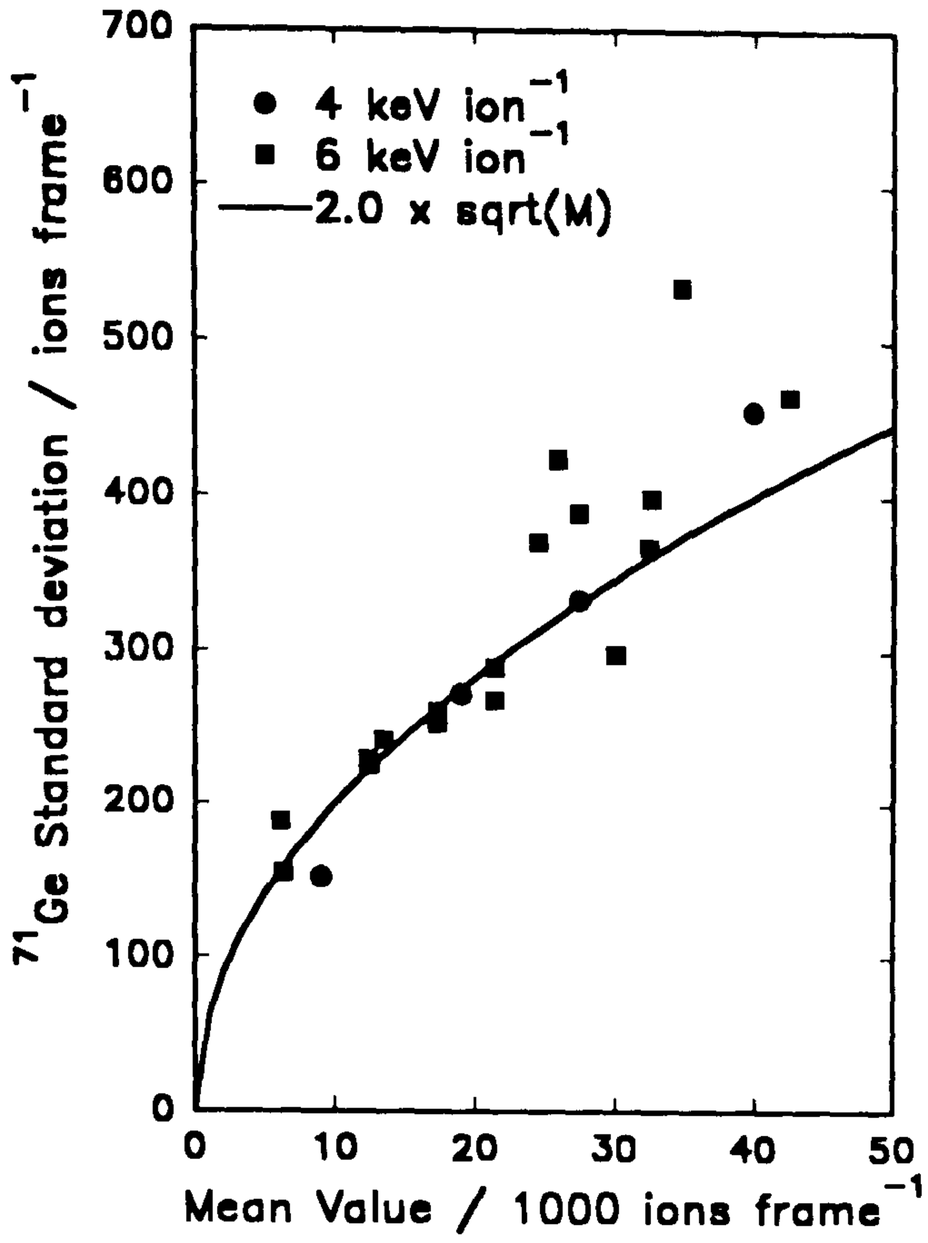


Figure 7.8. The relation predicted by the SiGe staircase.

The results for the germanium in the SiGe sample are shown in figure 7.8. Overlaid is a line approximately that from the other samples, $\sigma = 2\sqrt{m}$. The line is a good fit up to approximately 20000 ions frame^{-1}. Beyond this point the data show a large straggle and a linear dependence must be assumed.

This relation is superficially similar to the results from the preliminary study using a matrix signal (figure 7.3). Indeed, the germanium concentration in this sample did rise to 70%: far beyond the dilute limit. The reason for the apparently large difference in results between dilute and matrix-regime studies is unclear. One possible explanation is that forty thousand counts per ten seconds, with the gating set to approximately one sixteenth of the total frame time, implies a count rate above sixty thousand per second. It is known that the Channeltron detectors used saturate at count rates above about 1

MHz, and an ageing Channeltron may saturate at a lower frequency. This effect would reduce the higher count rates proportionately more than the lower frames - artificially reducing both the measured mean value and the standard deviation. The magnitude of this effect is difficult to estimate.

It was concluded that the distribution of noise on a dilute sample was, for our purposes, adequately represented by a Gaussian distribution of standard deviation $\sigma = 2\sqrt{m}$. This value was used to add noise to simulated depth profiles, and to calculate the expected deviation in the MaxEnt reconstruction process.

7.2. Appraisal of algorithms.

7.2.1. The histogram method.

The development of the histogram-based method was abandoned when it became apparent that the univariate search was significantly faster. At this point the method could optimise with up to thirty blocks, but was never very successful for data with a dynamic range of more than a single order. This failure was presumed to be due to the method of calculating step sizes, which forces each ordinate to approach the next asymptotically. This implies that any ordinate tending to zero will reach it only slowly.

7.2.2. The Univariate search.

The first test data set allowed the development of the univariate method to the state described in section 5.2. At this point the behaviour of the method with respect to the number of variables was assessed, with the results shown in figure 7.9 for a range of final step sizes. It should be remembered that smaller step sizes are advisable with this method. Tests

indicated that any reduction of the final step size below 10^6 atoms cm^{-2} would not significantly alter the reconstructed profile. It would, however, significantly increase the amount of time required to converge. Considering the results for a final step size of 10^6 atoms cm^{-2}, the number of convolutions required is just less than quadratic with the number of variables. Performing a reconstruction for data sets containing 256 points is thus feasible on a personal computer, with larger data sets feasible if left to run overnight.

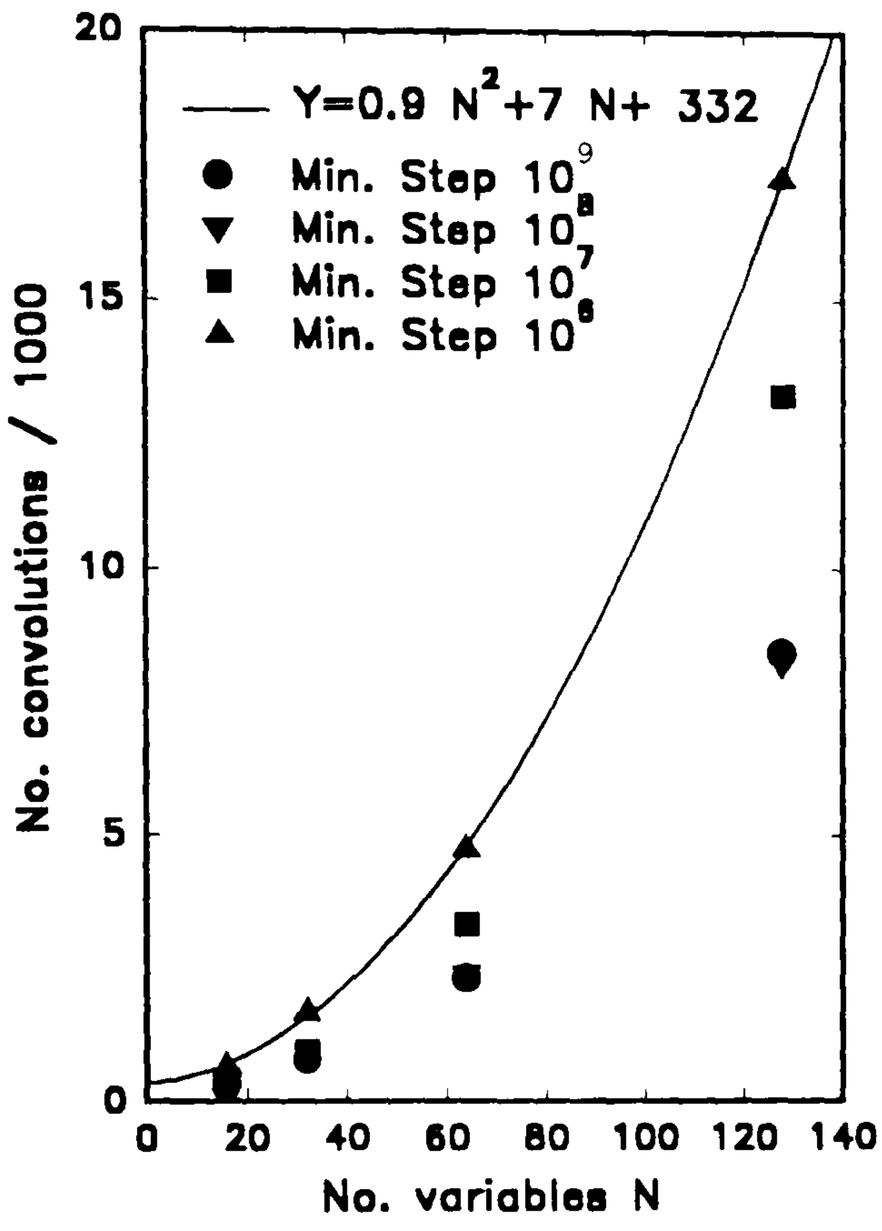


Figure 7.9. The speed as the number of variables is increased for the univariate algorithm.

7.2.3. The DSC algorithm.

For the Davies, Swann and Campey algorithm, the equivalent graph is shown in figure 7.10 for a variety of final step sizes. For this method it was determined that a final step size of 10^7 atoms cm^{-2} was sufficient. However, it can be seen that the automatic termination criterion do not punish the user for selecting a step size smaller than is required. Again, the behaviour is well described by a quadratic in *N*, but rather surprisingly the coefficients are

larger than those for the univariate method. Furthermore, the time required is much less well represented by the number of convolutions than for the univariate method, as each operation is a vector operation requiring time proportional to N. The Gram-Schmidt orthonormalisation process requires time proportional to N^2, which was ignored by this test. The algorithm is thus suitable for the reconstruction of the smaller simulated profiles. It has been found to be particularly useful as a test of the convergence of the univariate method, particularly regarding the final step size that should be used with that method.

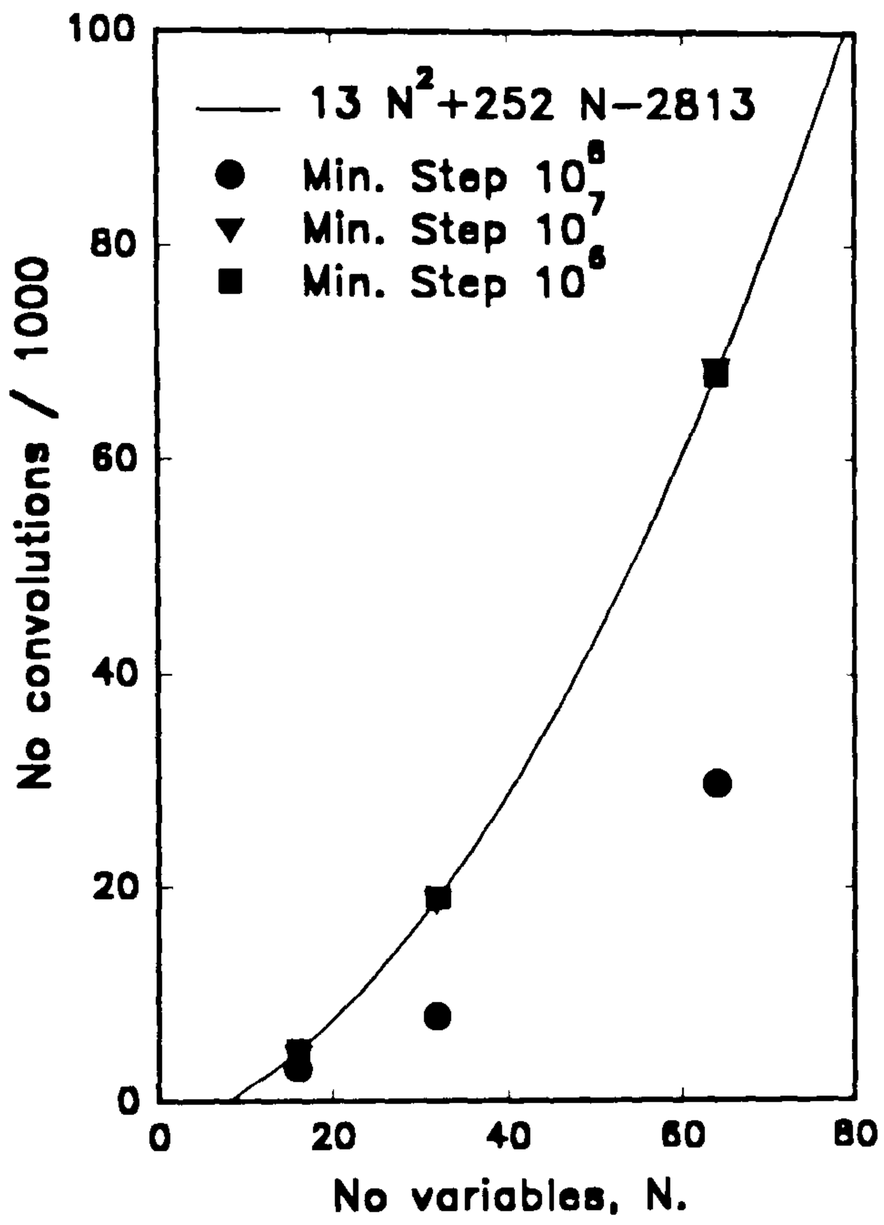


Figure 7.10. The speed as the number of variables is increased for the DSC algorithm.

One possible reason for the poor performance of the method is that the implementation of the non-negativity constraint was rather abrupt. This 'clipping' effect would cause the interpolation stage to be very inaccurate in some cases, as the actual difference between two consecutive steps could be

rather different from the step used. Attempts to improve the interpolation stage have not yet yielded any improvement.

Rather surprisingly, removal of the Gram-Schmidt process has been found to speed the optimisation process. This has not yet been explained, and no coding errors have been found.

In summary, development of the histogram method was discontinued, as it was significantly slower than either of the other two, while being the most complex and thus the hardest to develop. The univariate search is currently the most efficient, and all results given in this thesis will use this algorithm. The DSC method has been very useful as a test of the univariate search, but is too slow in practice to consider using at present.

7.3. Appraisal of MaxEnt Quantification.

In this section we present the results of the reconstruction of simulated depth profiles for the appraisal of the effectiveness of the MaxEnt method.

7.3.1. The reconstruction of depth-resolution parameters.

The test data of section 6.5 was simulated by convolving a measured response function with a concentration distribution that was designed to be simply parameterised. This data was intended to allow investigation of the region in which a significant improvement in accuracy could be obtained, the closeness of the MaxEnt result to the original profile, and some investigation of the relative performance of the MaxEnt method and Fourier deconvolution.

7.3.1.1. The full width at half maximum.

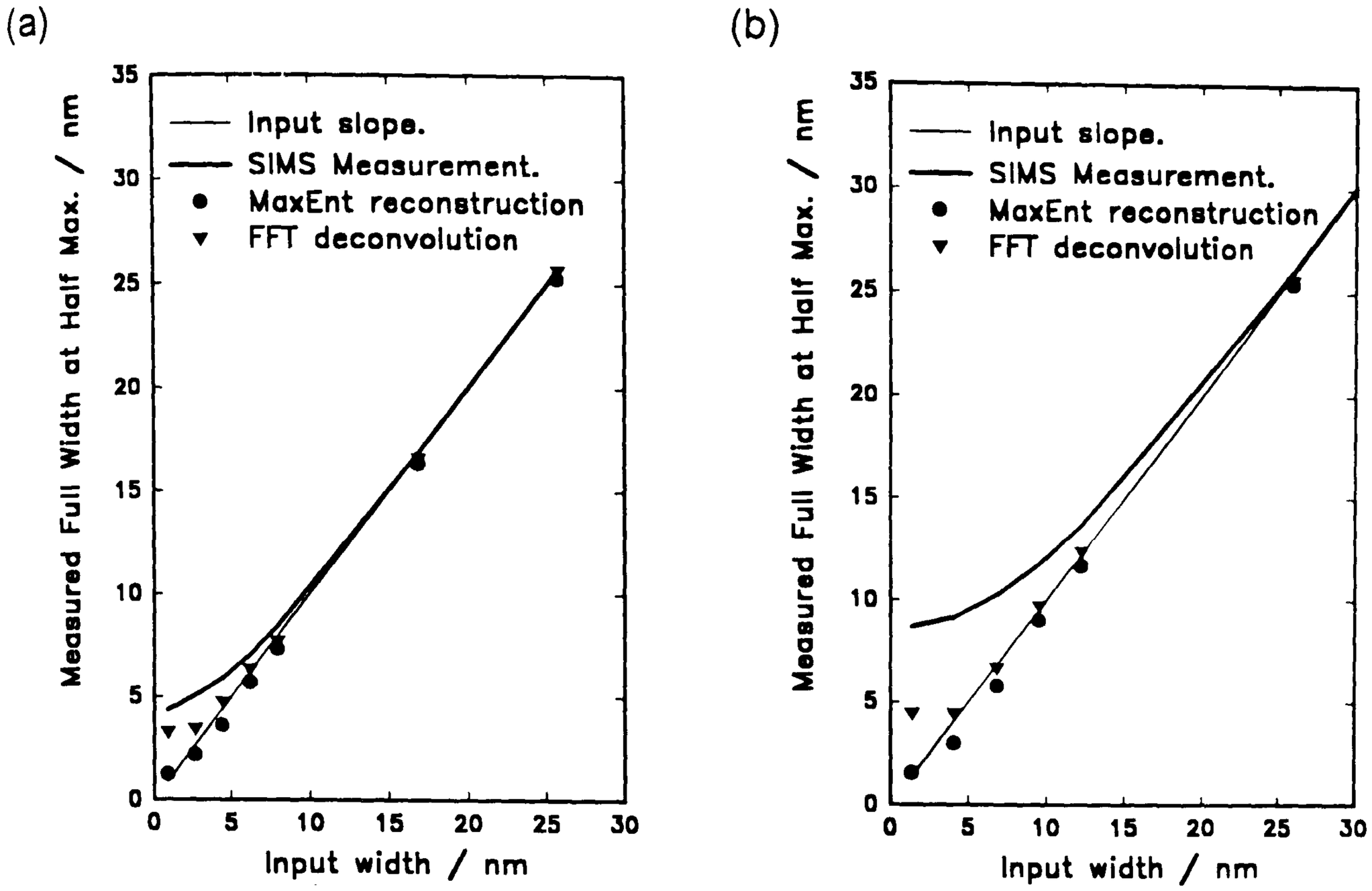


Figure 7.11. The FWHM parameter at (a) 2 and (b) 8 keV.

In figure 7.11 we show the results of the reconstruction of data simulated to investigate the reconstruction of feature widths, using the FWHM parameter. The simulated data were boron in silicon thin layers described in section 5.4.4.3. The FWHM was determined using the ProWin SIMS depth profile quantification software described in chapter 8. Results are shown only for 2- and 8- keV ion^{-1} bombardment; the investigations at 4-, and 6- keV ion^{-1} yielded similar results. The graphs show a number of features.

Regarding the accuracy of SIMS depth profiles using normal quantification methods (the thick line in figure 7.11). It is clear that for wide features (thicker than about 20 nm) the FWHM measured from the SIMS depth profile is a good estimate of the true value. However, for thinner features this accuracy is soon lost. Under 8 keV ion^{-1} bombardment the

thickness of a feature thinner than 20 nm wide would be overestimated. Reducing the primary ion energy to 2 keV ion^{-1} allows accurate measurement down to about 10 nm. At any normally accessible bombardment energy, the thickness of a layer thinner than this is grossly over-estimated.

Comparing the parameter from the MaxEnt reconstruction (filled circles) with the true value we see that the MaxEnt result is significantly more accurate than standard SIMS for thin layers. The slight non-linearity of the results for very thin features may indicate that a limiting thickness exists which even MaxEnt could not reconstruct. However, it was not possible to fabricate thinner profiles in these tests.

The results of the inverse FFT deconvolution method is also shown (filled triangles). It can be seen that in an intermediate range of thicknesses the FFT gives results that are better than standard quantification, but less accurate than MaxEnt. The smoothing was selected to just remove all negative values from the deconvolution, thus this result is the thinnest that FFT could produce: a more scrupulous user would have to use greater smoothing to reduce ringing, which was still evident in some profiles. It was also noted that the FFT result tended to have a smaller dynamic range than the MaxEnt.

7.3.1.2. The inverse slopes of a feature.

Moving to the reconstruction of inverse slopes at 8 keV ion^{-1}, using the test data described in section 6.5.4.4, the graph shown in figure 7.12 was obtained. The slopes were measured using the ProWin software, from 1% to 50% of the maximum value: it is a more error-prone calculation than the FWHM due to the presence of noise on the profiles.

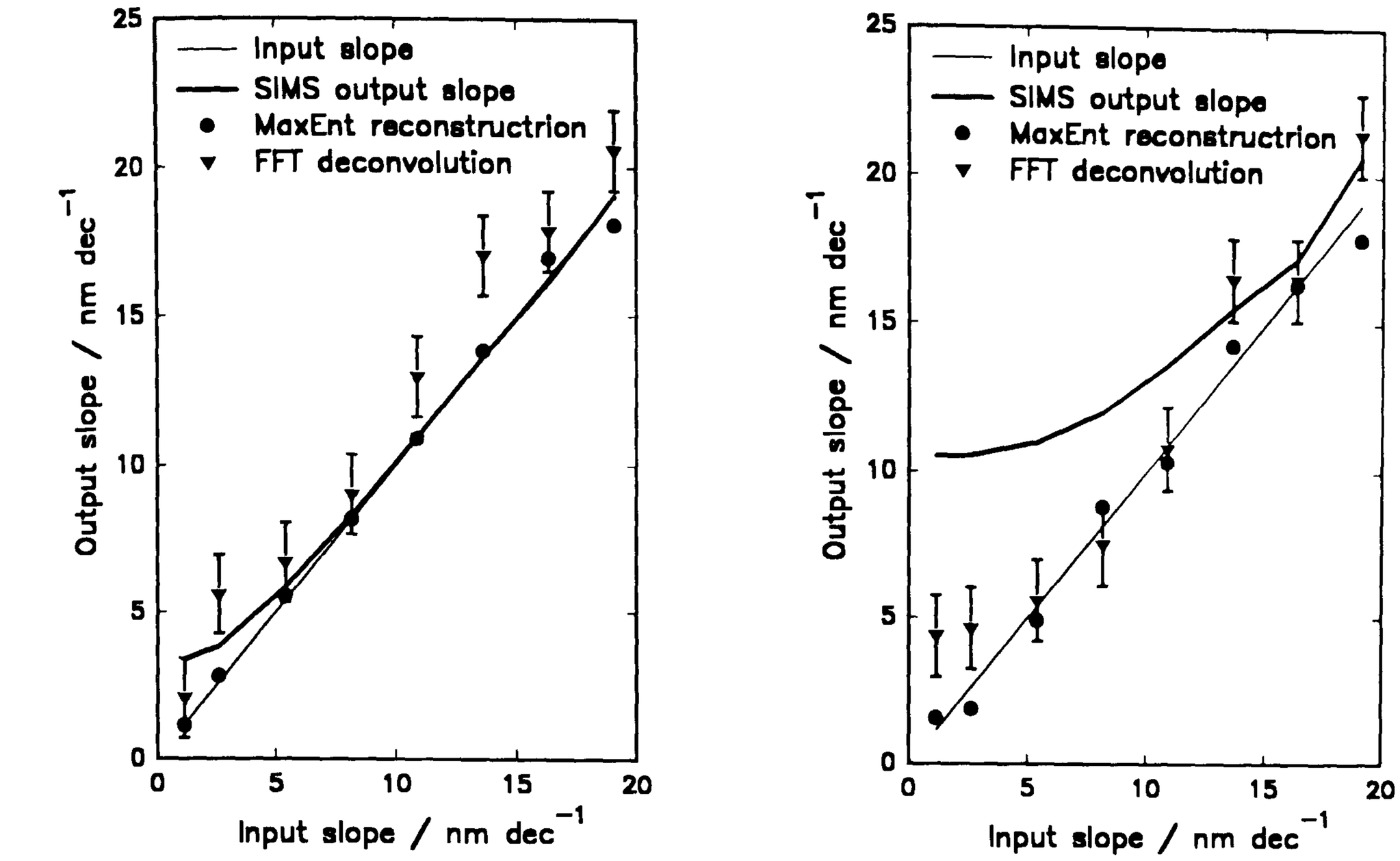


Figure 7.12. The reconstruction of the (inverse) slopes at 8 keV. (a) Rise slopes, and (b) Decay slopes.

Comparing the value as measured by SIMS (thick line) with the true value (thin line), it can be seen that the rise slopes as measured by SIMS are relatively accurate even under 8 keV bombardment. The decay slopes are, however, greatly overestimated - large room for improved accuracy exists.

The MaxEnt reconstruction (filled circles) gave a consistently more accurate estimate of the slope than the simulated SIMS profile. One interesting side-effect of this test was the realisation that the left-to-right nature of the univariate search made the algorithm faster for rise slopes than for decay slopes. However, tests indicated that the results of the reconstruction were not affected by the choice of direction.

For the FFT deconvolution, it was found that a relatively large amount of smoothing was required to produce data with an obvious slope, particularly

as the dynamic range tended to be reduced. For this reason the results are expected to be representative of the deconvolution technique rather than limiting. The remaining ringing on the features made accurate measurement of the slope rather difficult and errors were estimated at +/-1 nm decade^{-1}. Even allowing for this, the FFT results for rise inverse slopes are very poor, oscillating about the true value, and occasionally greater than would be measured by SIMS directly ! For the decay slopes the FFT was more reliable, but consistently gave results less accurate than the MaxEnt.

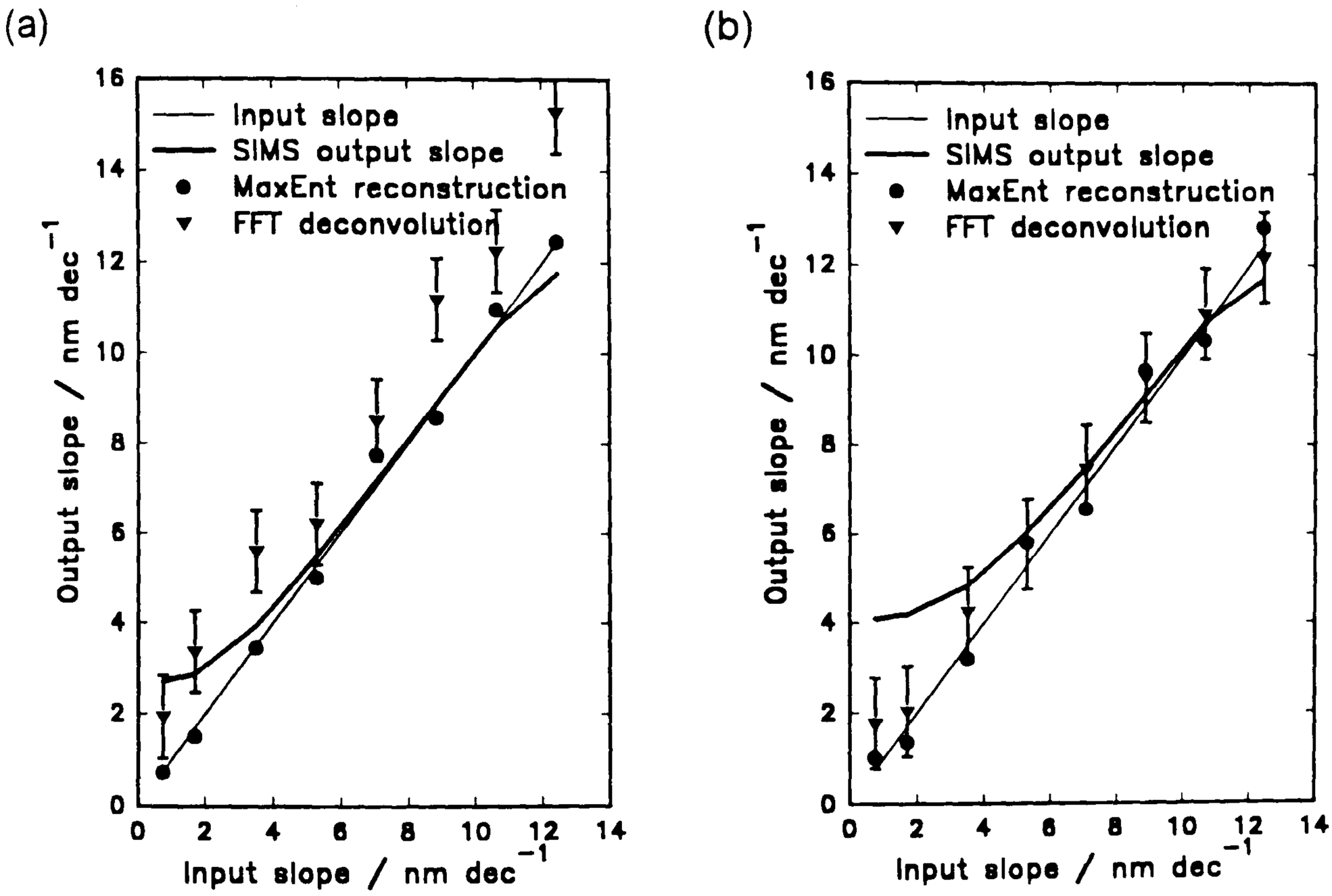


Figure 7.13. The reconstruction of inverse slopes at 2 keV, (a) Rise slopes, and (b) Decay slopes.

On reducing the probe energy to 2 keV ion^{-1}, the accuracy of the quantified SIMS profile is greatly improved. Here, rise slopes are accurate down to 4 nm decade^{-1}, and decay slopes down to 7 nm decade^{-1}. Within this (reduced) region of possible improvement, the MaxEnt reconstruction

follows the true value very well. Again, ringing on the FFT deconvolution reduces the usefulness of the technique significantly.

7.3.1.3. The reconstruction of peak heights.

Comparing the effects of convolution and reconstruction on the height of a peak under 2 keV ion^{-1} bombardment, the graphs shown in figure 7.17 were found. The smoothing effect of the entropy in the potential function can be clearly seen in figure 7.14a, where the low-concentration peak gave a simulated depth profile with low dynamic range. In figure 7.14b, where the dynamic range of the simulated depth profile was greater, the smoothing effect at the background level is far poorer. The smoothing used in the FFT method was more consistent, but again gave reduced dynamic ranges.

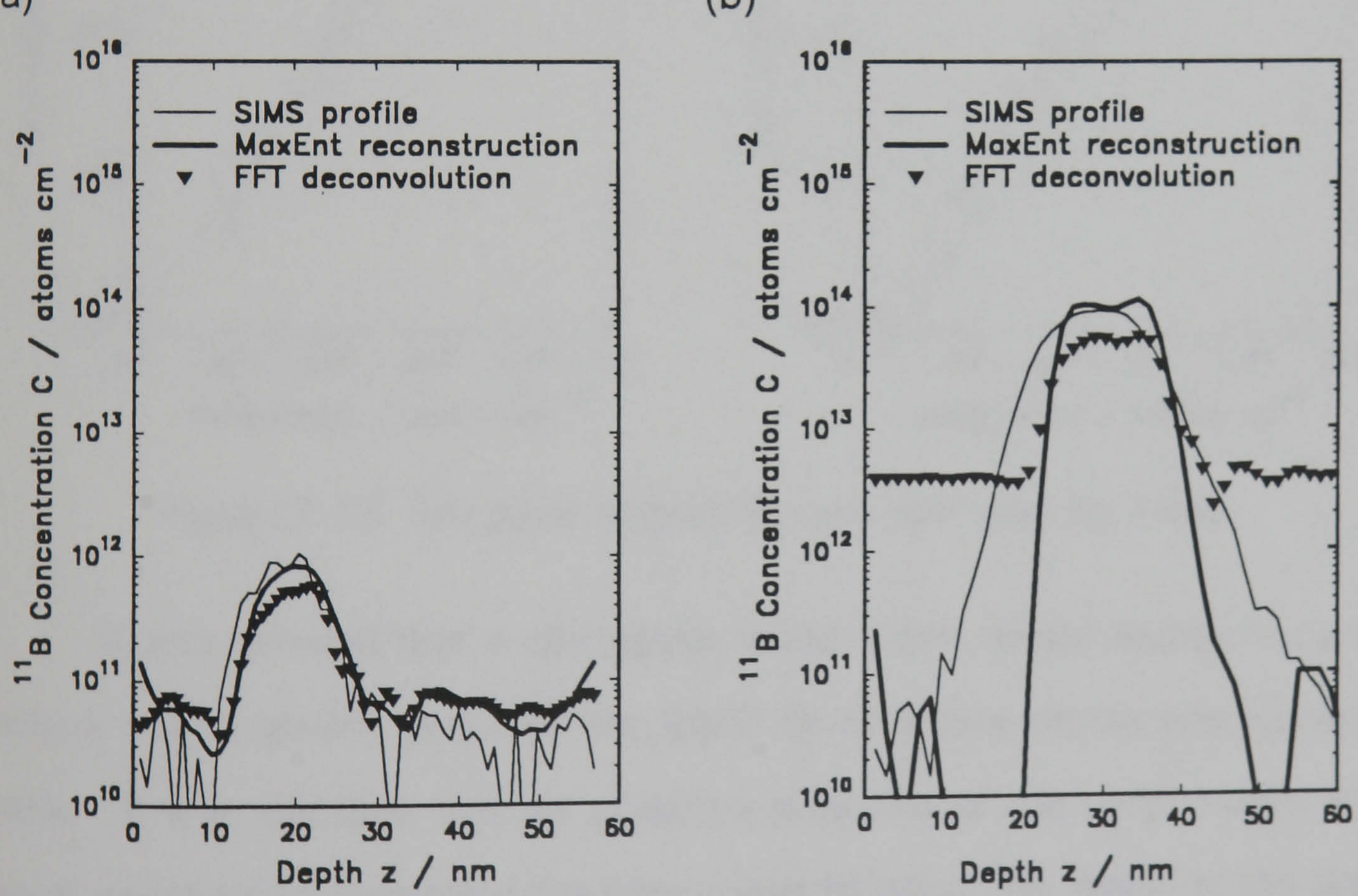


Figure 7.14. The reconstruction of peaks at 2 keV ion^{-1}, (a) the 10^{12} atoms cm^{-2} peak and (b), the 10^{14} atoms cm^{-2} peak (cf. figure 6.15).

A comparison of the peak concentration obtained by all three quantification procedures was difficult: for the SIMS data, the broadening of the feature gave an asymmetric peak shape rather than a flat distribution; the FFT deconvolution resulted in ringing, making the definition of peak a little ambiguous; for the MaxEnt reconstruction, slight variations from a flat profile also existed.

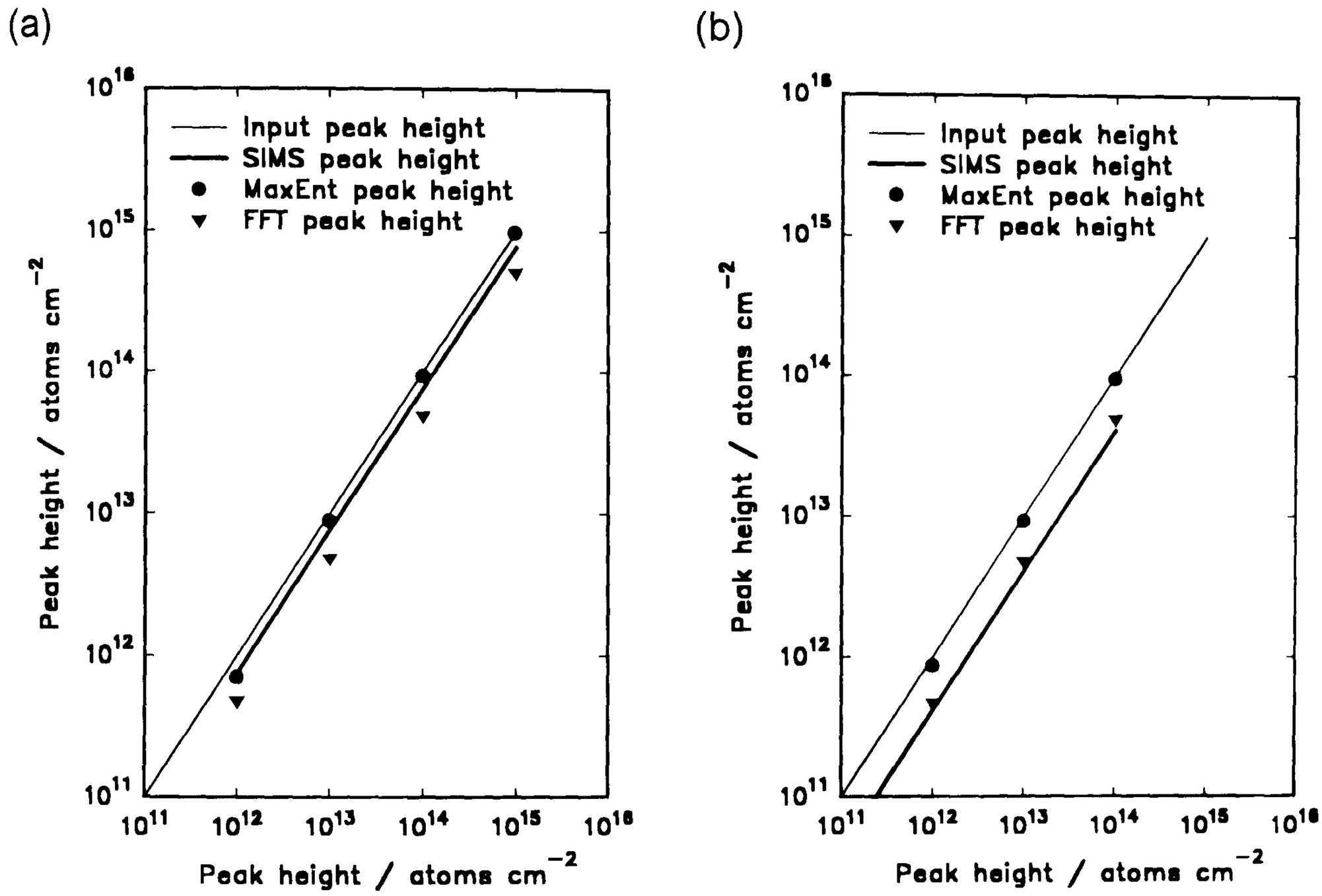


Figure 7.15. The peak heights at (a) 2 keV, and (b) 8 keV.

It was decided that a calculation of the mean height across the entire feature would greatly penalise the SIMS data, as the profile was generally wider. It was decided that an objective measure of the peak height in all cases would be to determine the mean peak height *in the region of the actual concentration profile only*. This comparison is shown for 2- and 8- keV ion^{-1} bombardment in figure 7.15. As expected the blurring effects of the atomic motions cause the peak height in SIMS to be artificially low, with a greater

effect at higher probe energy. The MaxEnt reconstruction gives a mean peak height significantly closer to the true value. The FFT deconvolution, due to the smcothing required, was better at 8 keV but actually *worse* at 2 keV than standard SIMS quantification.

7.3.2. The reconstruction of an ideal delta layer.

In figure 7.16 we show the results of the reconstruction of the delta layer in ideal circumstances (section 6.5.5). It is clear that the effect of the entropy has been to predict a feature with a finite width. The agreement between the different energies is exceptionally good (c.f. conventional profiles in figure 6.12a), with almost identical widths, positions and heights.

The results of FFT deconvolution of the same sample are shown in figure 7.17. Smoothing was again minimal, to remove only negative values; ringing is clearly visible at 6- and 8- keV. The profiles do represent an improvement over standard SIMS quantification (figure 6.12a) in some respects - the profiles are symmetric and show no differential shift. However, a loss of dynamic range is apparent.

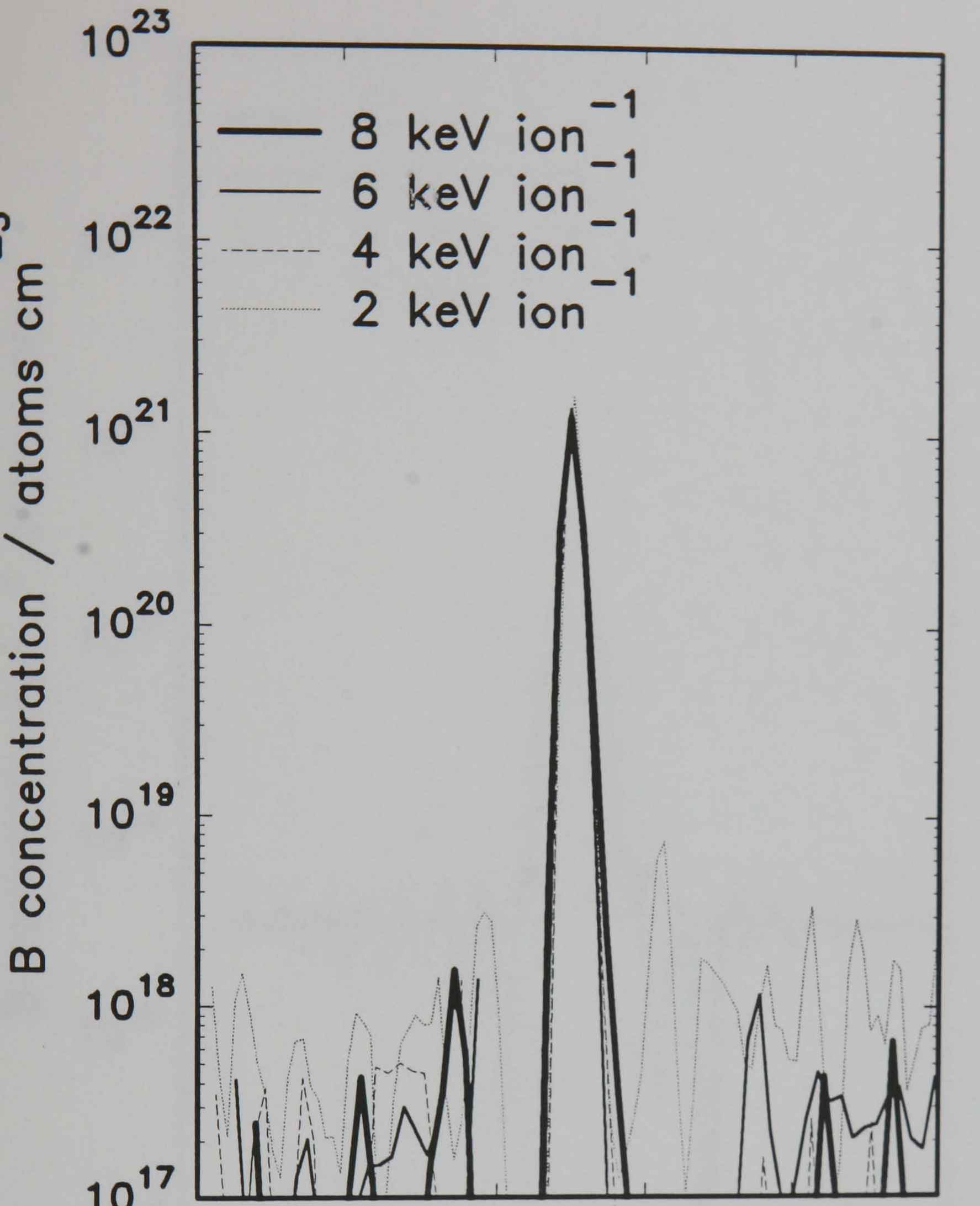


Figure 7.16. The reconstruction of a true delta layer in ideal circumstances.

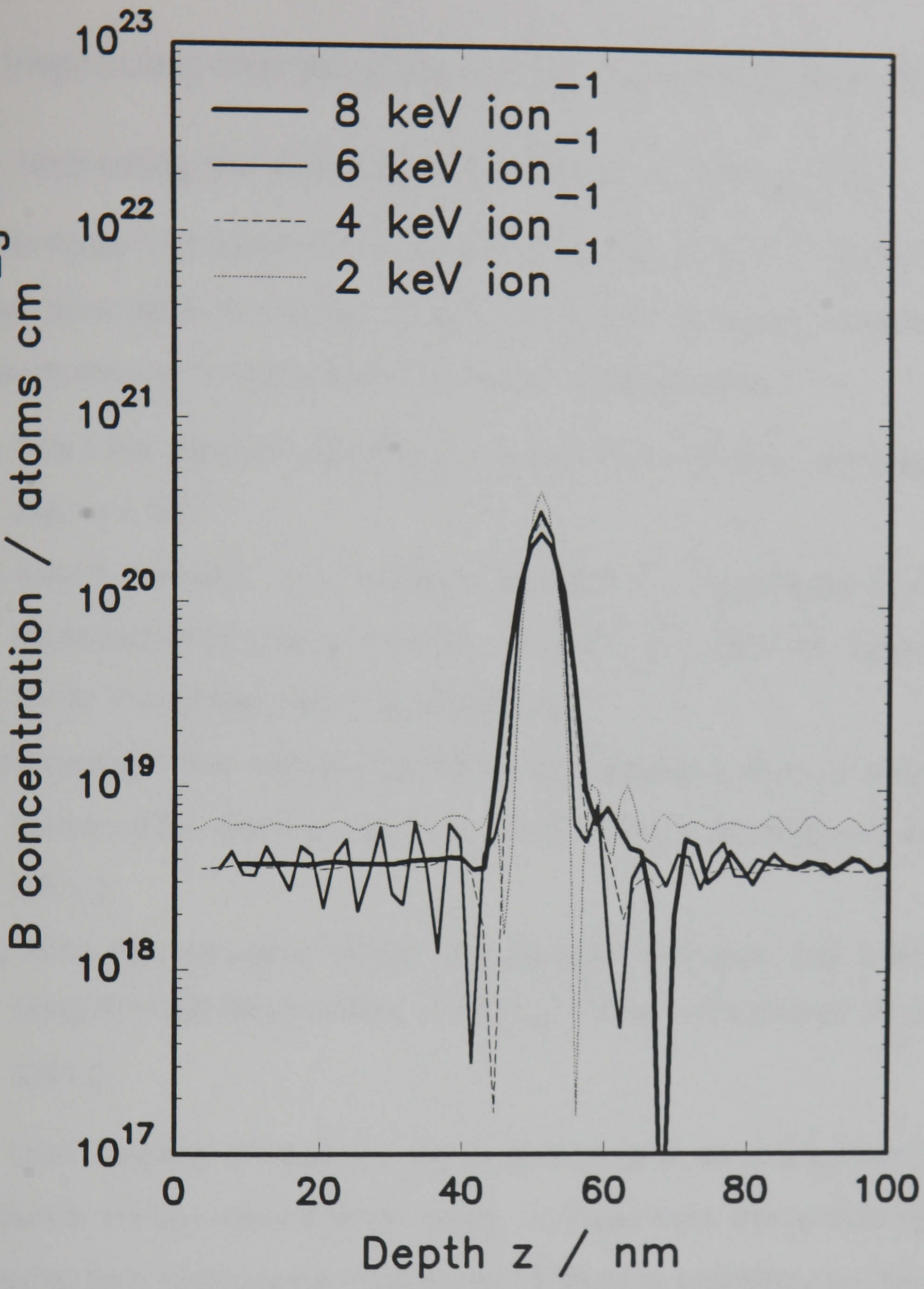


Figure 7.17. FFT deconvolution of the true delta.

7.4. Improving the background on the reconstruction.

7.4.1. Improving the estimate of the noise on the profile.

In figure 7.18 we show the results of reconstruction of simulated depth profiles (described in section 6.4.2) using the following methods for calculating the standard deviation σ from the measured data Y

(i) Basic: the standard deviation is calculated directly from the measured data, $\sigma = 2\sqrt{Y}$.

(ii) Added constant: the standard deviation is calculated from the measured data plus a constant, $\sigma = 2\sqrt{Y + b}$, where for these tests $b=10$. This is the method of section 6.6.1.1.

(iii) Smoothed: the standard deviation is calculated from a smoothed version of the depth profile Y', $\sigma = 2\sqrt{Y'}$. This is the method of section 6.6.1.2.

(iv) Using the calculated profile: the standard deviation was calculated using the calculated profile, $\sigma = 2\sqrt{Y_{calc}}$. This is the method of section 6.6.1.3.

Both methods (ii) and (iii) were found to give an improvement over method (i) in the low-concentration region. In these tests, the reconstructions obtained by both methods were too similar to allow an unambiguous choice of the 'best' method. Method (ii) was found to result in slightly faster convergence than method (iii). Also, as was stated earlier, the smoothing used in method (iii) may potentially *reduce* the calculated noise deviation on sharp peaks. It was thus decided that method (ii) would be the more general-purpose method of improving the estimate of σ.

Method (iv) was found to be particularly disappointing. It had been thought that, as the simulated depth profile is (by definition) noise-free, the calculated noise deviation would be smoother in the background region. However, the simulated depth profiles were found to follow the 'measured' profiles even at the noisy low-intensity level. This effect is due to the very 'thin' nature of the SIMS response function - the peak of the response at 4keV has typically only five or six data points. In the authors opinion the use of response functions with more data points will reduce the ability of the simulated profile to follow the noise on the measured profile. At this point the method may give considerable improvement. The method does, however, slow the convergence as expected (requiring up to 50% more convolutions) and removes the possibility of pre-computing σ.

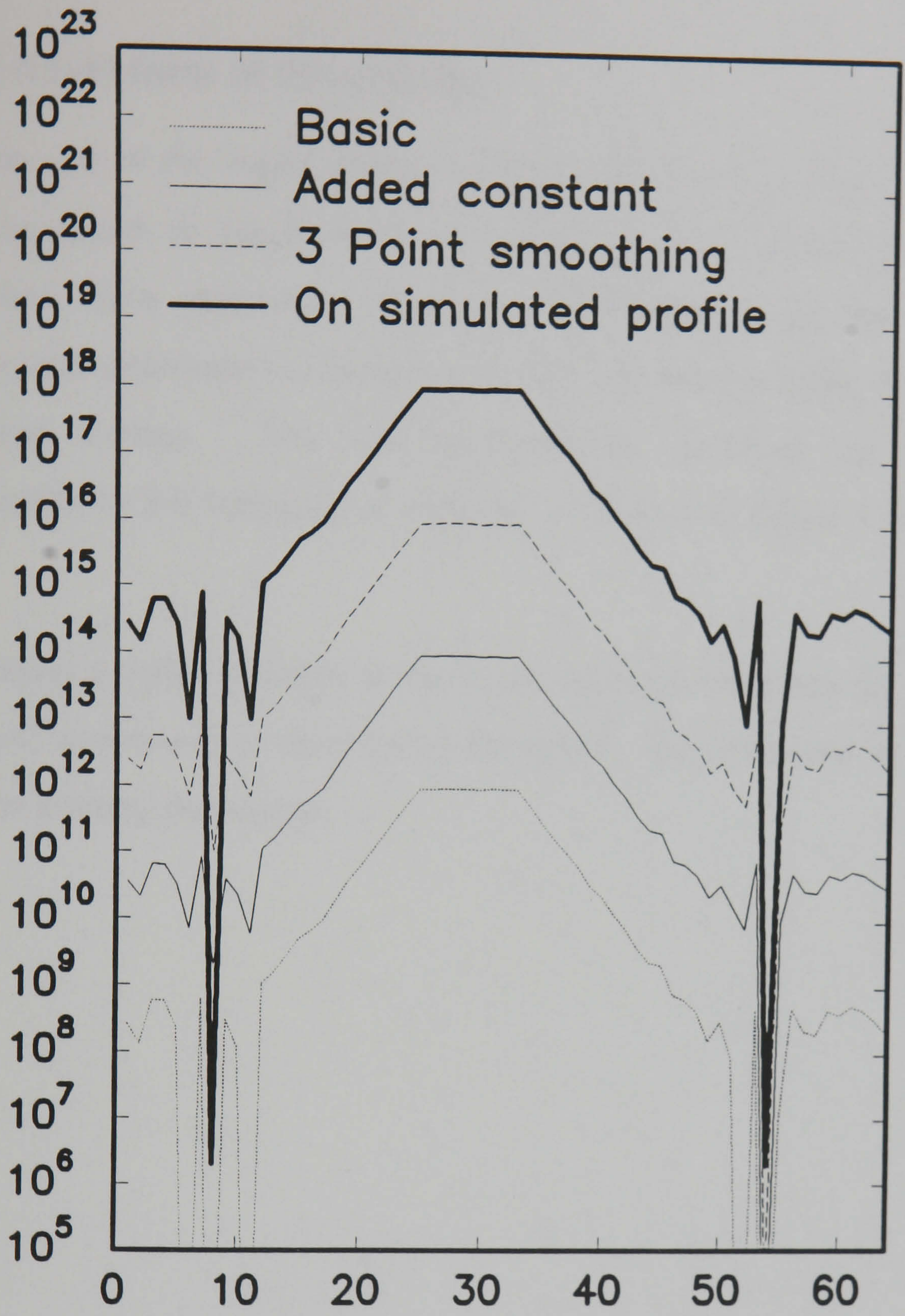


Figure 7.18. The effect of changing the estimate of the standard deviation on the reconstruction of a single peak, using the Shannon-Jaynes entropy. The profiles have been offset vertically for clarity.

7.4.2. A novel form of the entropy.

The use of the logarithmically compressed entropy function (section 6.6.2) was found to result in slower convergence (typically 50% more convolutions were required). This is understandable, as the potential function in the optimisation is significantly less well-behaved than when using the normal entropy. This was accompanied, however, by improved reconstruction in the background regions, as shown in figure 7.19 (dotted line).

Adding a better estimate of the noise deviation (methods (ii), (iii), and (iv), above) was found to have very little effect - the behaviour of the new form of the entropy dominates.

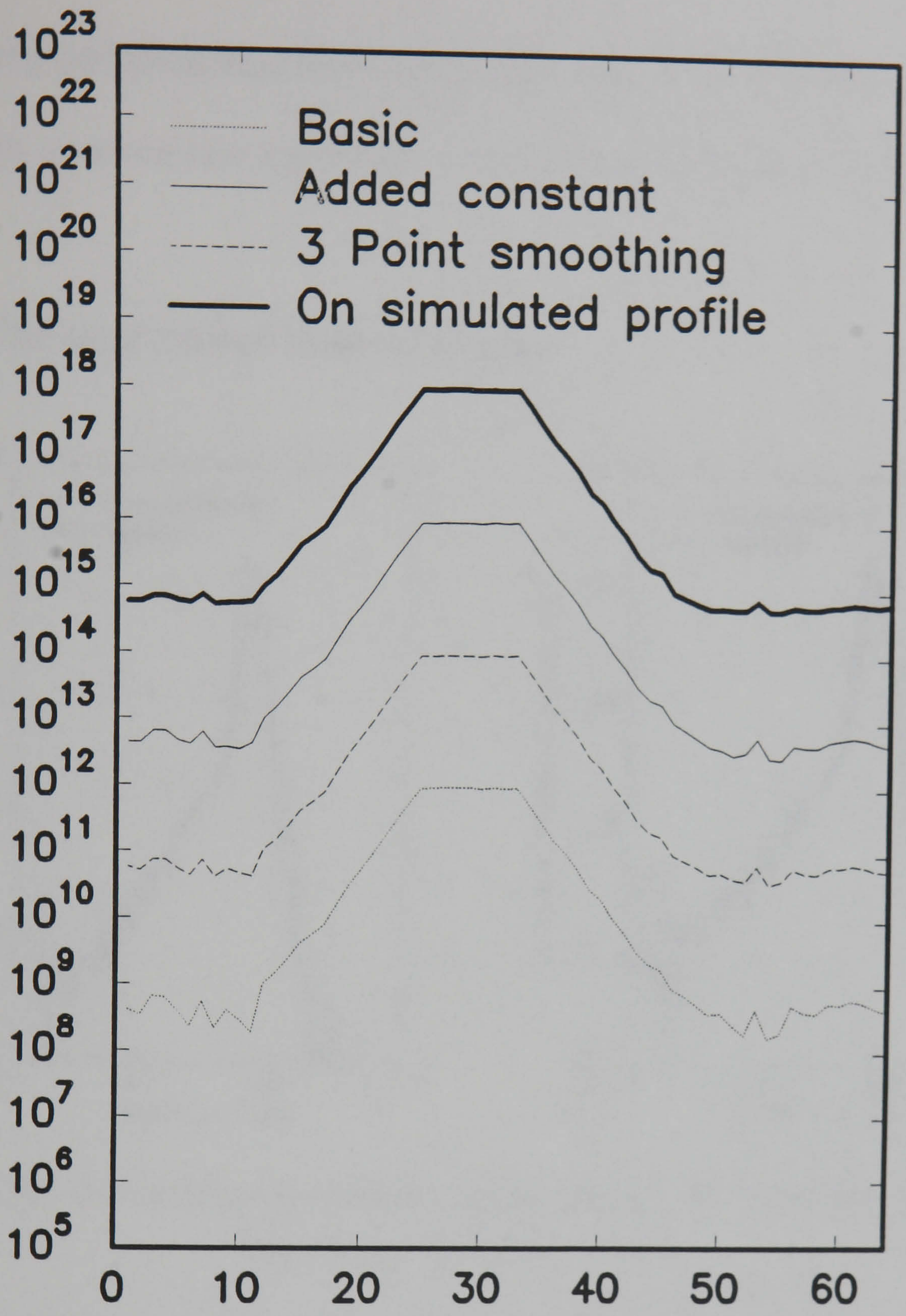


Figure 7.19. The effect of changing the estimate of the noise deviation while using the logarithmically-compressed entropy function. The profiles have again been offset vertically for clarity.

7.5. The reconstruction of SIMS depth profiles.

We describe now the results of reconstruction of genuine SIMS depth profiles.

7.5.1. The segregated GaAs(Si) layer.

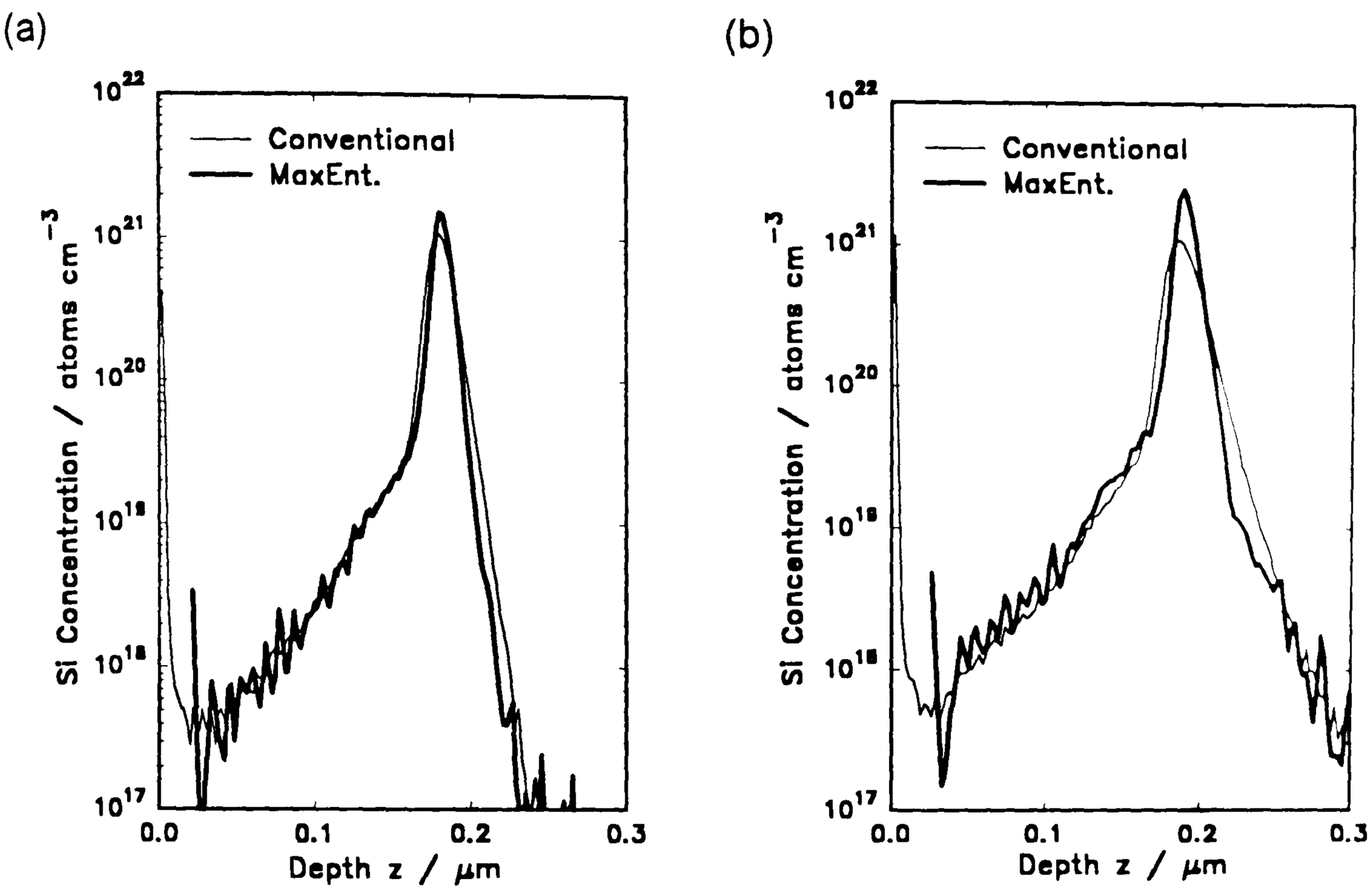


Figure 7.20. Conventional and MaxEnt quantification for the broader GaAs(Si) layer at (a) 6, and (b) 15 keV.

In figure 7.20 we compare the results of MaxEnt quantification with those from conventional quantification for the 0.7 nm GaAs(Si) layer. At both bombardment energies the MaxEnt reconstruction is clearly sharper, and shifted slightly to the right, compensating for the shifting effect of the asymmetric response.

Closer examination revealed a noticeable different in peak height between the two MaxEnt reconstructions, and for the 15 keV profile there was

a significant difference in total (integrated) Si concentration between the MaxEnt and conventional SIMS quantification. This discrepancy has been explained by the behaviour of the reference channel at 15 keV: during the measurement of the delta layer the As^- channel used as a reference had a mean value of 661 ions $frame^{-1}$, while during the measurement of the segregating sample, the value was 849 ions $frame^{-1}$. If an equivalent rise in signal for the measured $AsSi^-$ ions is assumed (this is only a first approximation) a correction may be performed: we correct the predicted MaxEnt result by multiplying by 0.78. This gave much better agreement between MaxEnt and conventional SIMS quantification with regard to total impurity concentration. Indeed, the agreement at both bombardment energies was greatly improved.

A difference remains, however, in the position of the peak of the two MaxEnt reconstructions. As this difference is also present in the normal data these errors must be assumed to be caused by an inaccurate depth scaling.

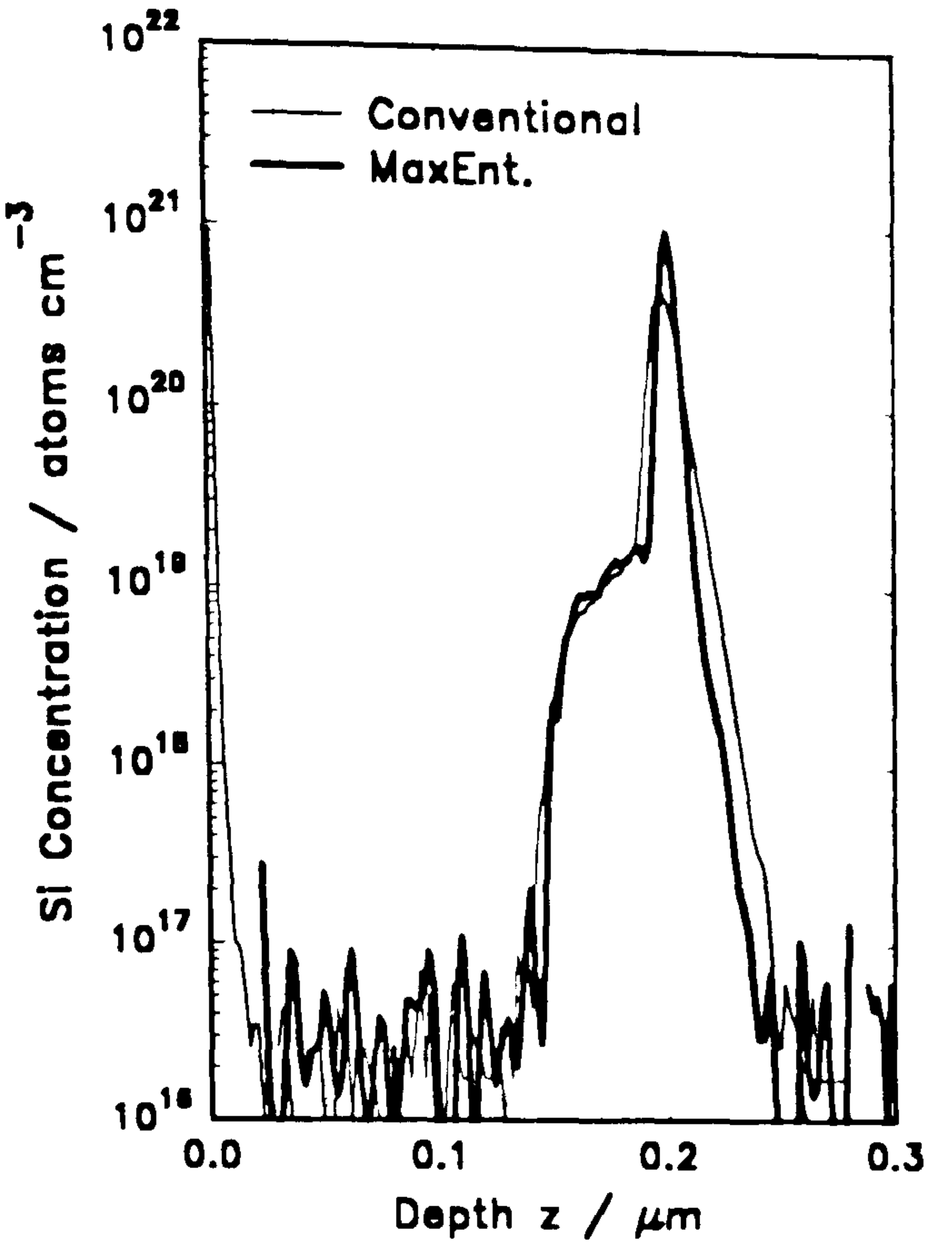


Figure 7.21. Comparison of MaxEnt and conventional SIMS quantification for the 0.1 nm GaAs(Si) thin layer.

The reconstruction of the thinner Si layer (figure 7.21) is less informative, as no comparison exists; it is included for completeness. Again, improved resolution is obtained.

7.5.2. The Si(B) triple delta.

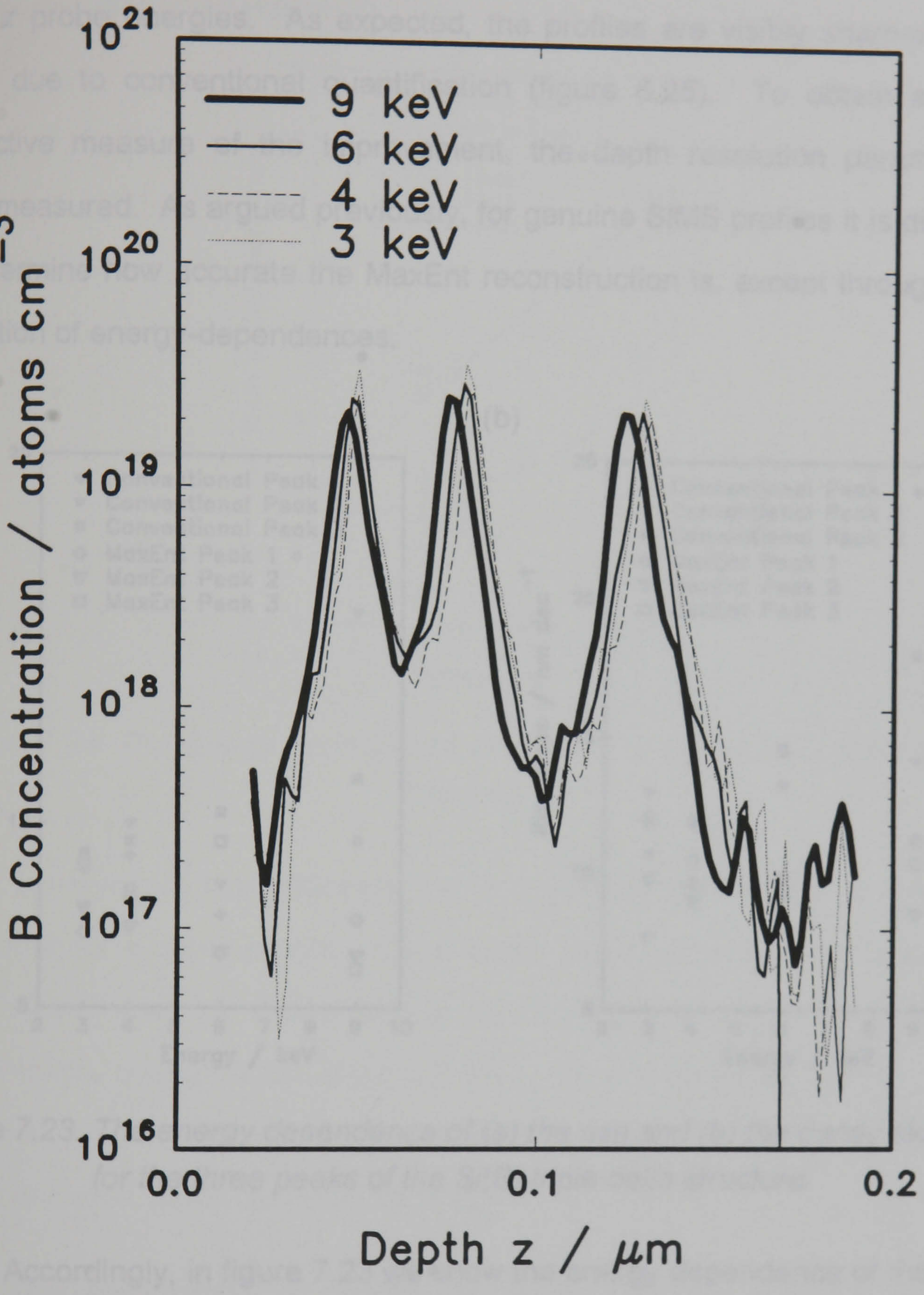


Figure 7.22. MaxEnt quantification of the Si(B) triple delta structure at a range of primary ion energies.

Shown in figure 7.22 are the results of the MaxEnt reconstruction for all four probe energies. As expected, the profiles are visibly sharper than those due to conventional quantification (figure 6.25). To obtain a less subjective measure of the improvement, the depth resolution parameters were measured. As argued previously, for genuine SIMS profiles it is difficult to determine how accurate the MaxEnt reconstruction is, except through the reduction of energy-dependences.

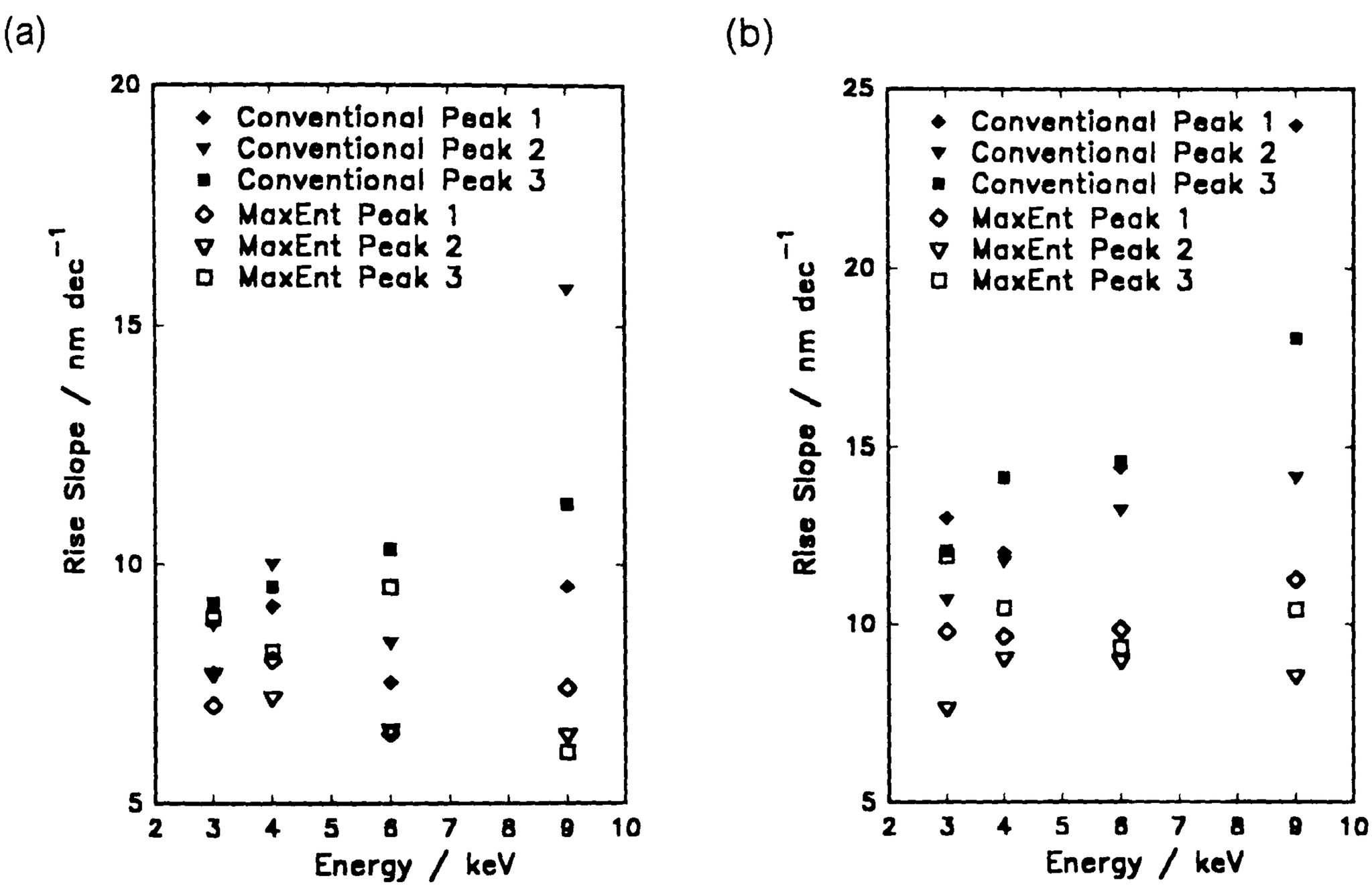


Figure 7.23. The energy dependence of (a) the rise and (b) the decay slopes, for the three peaks of the Si(B) triple delta structure.

Accordingly, in figure 7.23 we show the energy dependence of the rise and decay (inverse) slopes for both conventional and MaxEnt quantification. For the conventional data, the rise slope showed a small energy dependence, and the decay slope a greater dependence. These effects are characteristic of the boron in silicon system. However, for the MaxEnt quantified data

neither parameter had an energy dependence to within the scatter of the data: a very promising indication.

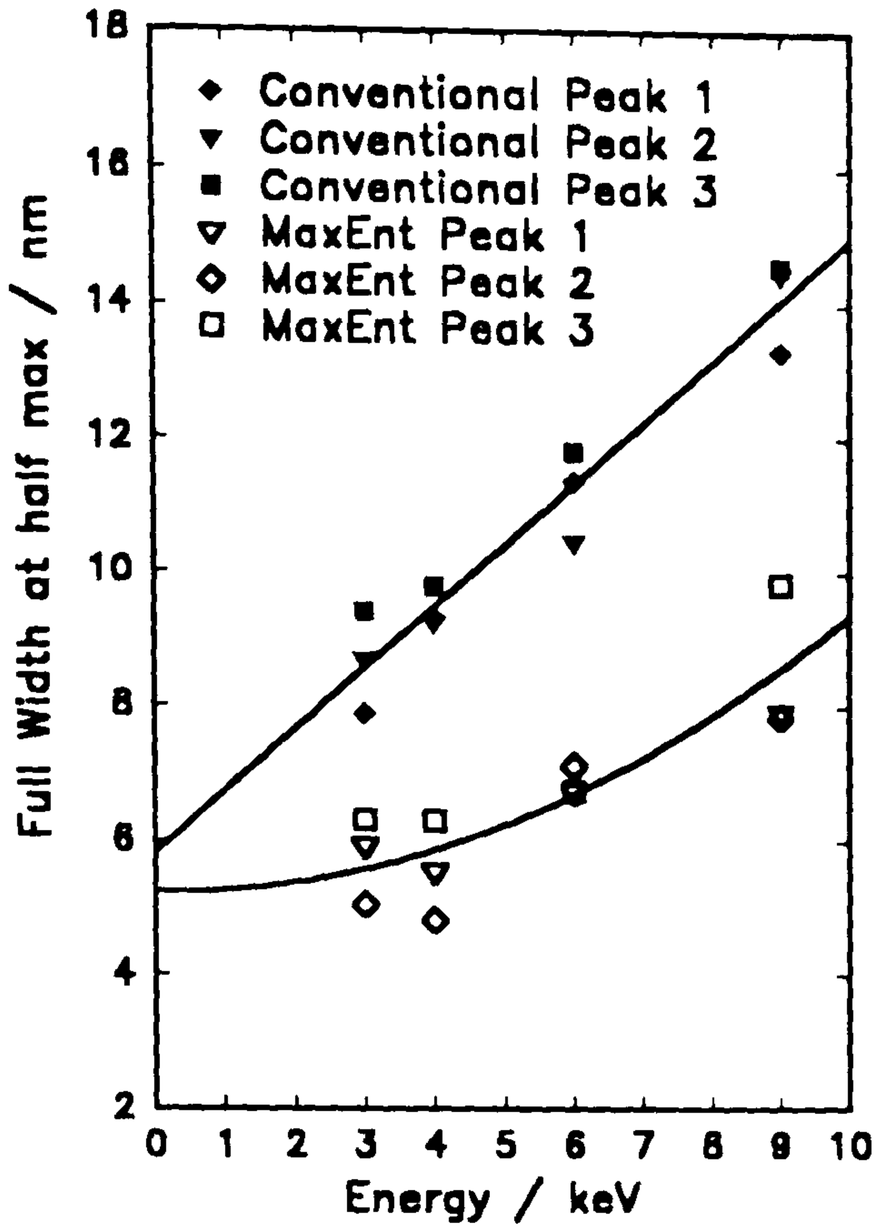


Figure 7.24. The energy dependence of the FWHM for the Si(B) triple delta structure.

Experiments had indicated that the FWHM was a less error-prone parameter for a depth profile. It was found that the MaxEnt reconstruction had reduced the full widths almost to one half of the value given by conventional quantification (figure 7.24). At low energies, there is an indication that the value reached becomes energy independent, while at higher energies, some dependence remains.

The appearance of energy independent feature parameters is a reasonable indication that the 'true' values of these parameters have been recovered by the MaxEnt reconstruction. For this sample, the predicted (inverse) slopes and widths are too large for the layers to be considered delta layers. Accordingly other ways were sought to determine the thickness of the layers. The most simple was the particularly effective: figure 7.25 shows the

third peak of the triple delta structure, overlaid with the peak of the delta layer used to measure the response, both are raw SIMS profiles under 6 keV bombardment. The difference in the size of the features is very clear. We may conclude that the triple 'delta' is not, in fact, as thin as was hoped. The increased width may be due to the extra growth steps required for the production of the two other delta layers on top.

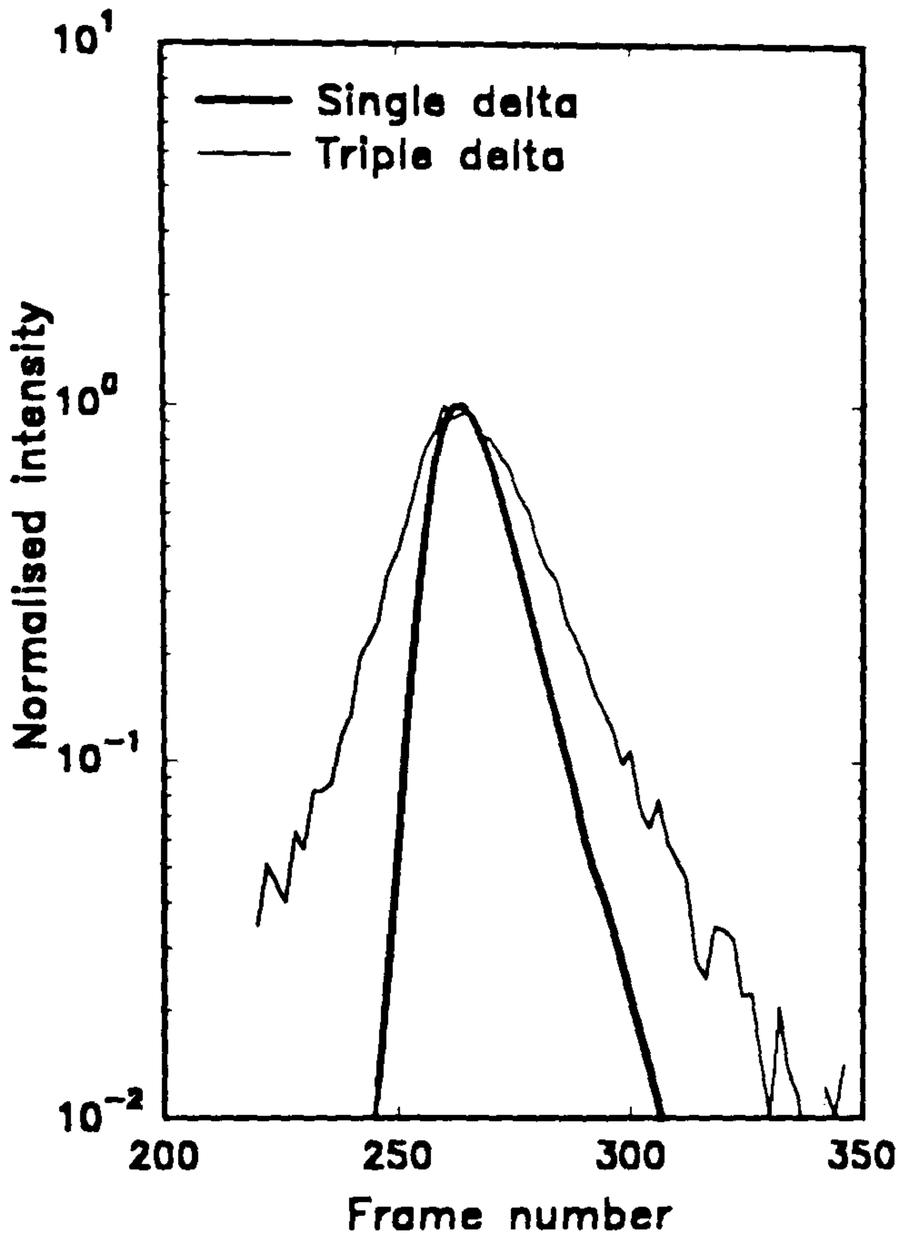


Figure 7.25. The single and triple Si(B) deltas, shown profiled at 6 keV.

There exists an alternative explanation, however, that must be considered. The slight energy dependence of the full width of the MaxEnt reconstructed profiles above the primary ion energy of 4 keV ion^{-1} may indicate that the 'true' profile has not, in fact, been obtained (at least at these energies). If this is the case we should determine why.

(i) It is to be expected, that at higher energies, the greater information loss will limit the amount of information that can be obtained, thus we expect reconstruction to yield less informative profiles.

(ii) Alternatively, the response measured may have been inappropriate. If the samples differ in an unmeasured manner such that a different mix of atomic motion process occur, the response in the two samples need not be the same. The idea that samples of different material 'quality' have different response functions is supported by a survey of decay slopes in the literature: different magnitudes *and* different energy dependencies have been obtained by different authors, with the sharpest profiles and least energy-dependence for the most recently grown samples (Dowsett and Barlow, in press). This rather disturbing conclusion is, however, far from fully characterised, and may be inappropriate here, as the delta layer and triple-layer sample were both grown concurrently, under identical conditions in the same instrument: no large variation in material 'quality' is to be expected.

8.0. The quantification software ProWin.

Several processes are required to fully quantify a depth profile, obtaining concentration versus depth information from the raw data. The conventional methods were detailed in chapter 2. These calculations are relatively simple, but tedious to perform manually, thus the SIMS group at Warwick makes great use of in-house depth profile quantification software. The main role of the software is to interpret the data file and then to allow various transformations of the data, either in interleaved form (depth calibration) or as separate channels (for the calculation and application of sensitivity factors. The quantified depth profile may then be stored as a file or as hardcopy.

8.1. The original quantification software.

Development of quantification software was originally done in RM-BASIC on an RM-380Z (Dowsett, Heal, Fox, and Parker, 1985), just as the control software was itself developed. The quantification software was accompanied by an RM-BASIC hardcopy program to obtain output on a Hewlett-Packard six-pen plotter. Both of these programs had been ported to RM-compatible PCs in the early eighties. Later, a BASIC program was added to calculate the 'depth resolution' parameters.

8.2. The 'PROED' quantification suite.

Much of the quantification software had been replaced in the mid-eighties by an IBM-compatible PC version. The name was derived from 'depth PROfile EDitor', and was written in FORTRAN 77. The functionality was provided by a suite of executable programs, selected by a controlling user-interface (menu) program. Each (sub-) program accessed the depth

profile data stored in random-access files on the hard disk. The data were stored as X-Y pairs, still in interleaved format. The PROED suite had a number of shortcomings:

- Obtaining hard-copies remained a two-stage process: the FORTRAN code created a file that could be used by the original RM-BASIC hardcopy program. This was in itself rather restrictive, as RM-BASIC is poorly supported on other PCs. Also, the format of these graphs was not in a publishable form, requiring the user to enter the data into a commercial software package to manually reformat the graphs.
- The calculation of the 'depth resolution' parameters often quoted in the literature also required transfer to the older BASIC program. This did not require a change of PC, as this program was PC-compatible. It remained, however, rather irritating.
- Many of the quantification stages required the user to write down a number (e.g. a sensitivity factor) only to re-enter it at a later stage. This was due to the implementation as a suite of separate executables with limited communication, and potentially allowed user-error to creep into the quantification process.
- Although SIMS quantification had developed much since the original RM-BASIC software was written, the PROED suite still allowed arbitrary (i.e. meaningless) transformations of the depth profile data. It also lacked some of the more complex transformations, such as those required for the depth quantification of multi-layer structures, which had been present in the original software.
- The continued use of interleaved data during quantification limits the software to depth profiles with a single region of interest (termed a *gate*). Many modern instruments (including the EVA3000 instrument at

Warwick) have the capability to define multiple gates and it is vital that these depth profiles can also be quantified accurately.

8.3. The design of the new software.

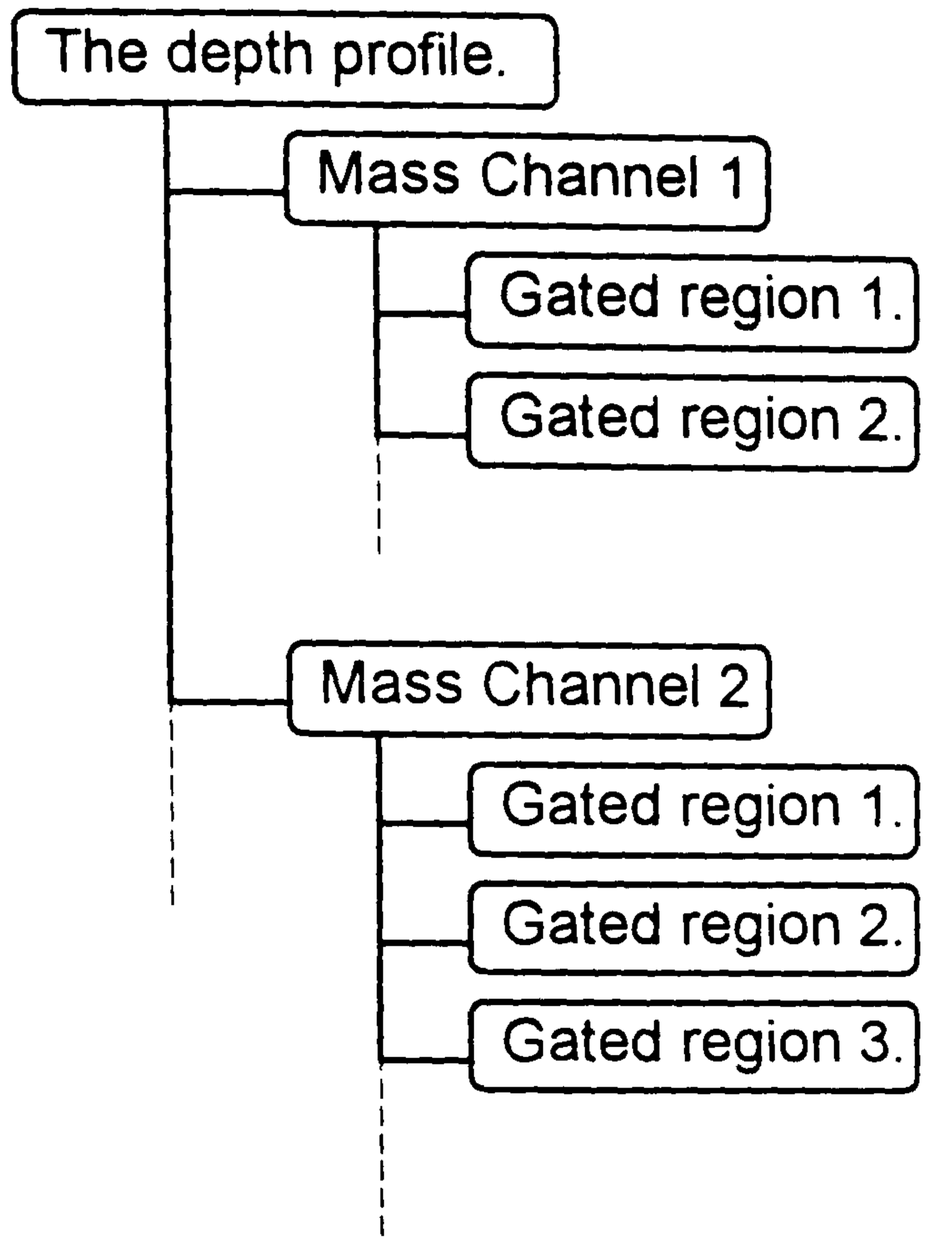


Figure 8.1. The data structure used in the new quantification software.

The major stage in the design of the new software was the specification of the new data-structure. To cope with depth profiles from instruments of ever-increasing flexibility, the data should be stored in an equally flexible manner. To this end, the structure chosen was an *n-tree*, as shown in figure 8.1. In order that the number of branches is not fixed, dynamic memory allocation techniques were required rather than fixed size arrays. The structure should also remember a number of forms of the data: frame numbers, primary ion dose, or depth, may be plotted against signal intensity (counts) or concentration. This would allow the user to return to the raw data at any stage.

The second stage of the design process was to redesign the format of the output, so it would be immediately acceptable for inclusion in research journals. A standard page size (A4) was selected, and the size and position of the graph, the key, and the figure caption were chosen to give an attractive graph on a page with adequate (one and a half inch) margins all round. It was decided that the program should provide true WYSIWYG (What You See Is What You Get) data processing, so the user can always see the current state of the data as it will appear on the hardcopy.

8.4. Implementation.

The software was implemented using object-oriented C++. The language was chosen because C provides the tools necessary for the use of dynamic memory allocation, and object-oriented coding can massively reduce the possibility of coding errors. The program was written to run under Microsoft Windows, giving the graphical tools and the device-independence required for WYSIWYG, plus support for many types of hardcopy. This choice also ensures a number of other advantages such as a standardised user interface and the possibility of Cut-and-Paste, such that the graph of a fully quantified depth profile can be placed directly in a word processor document, removing the need for scissors and glue.

At time of writing, version 1.0 of the new quantification software has been in use for eighteen months, during which time comments from various users have continued its development. All the functionality of the original software has been included. Version 2.0 of the software has just been released. This version has been developed to give far better support for multiple-gate profiles (although no such profiles have yet been performed).

9.0. Conclusion.

9.1. MaxEnt quantification of SIMS depth profiles.

The objective of this research was to investigate the possibility of obtaining increased accuracy in SIMS depth profiling by the use of improved quantification methods, specifically Maximum Entropy (MaxEnt) reconstruction. This was considered the first stage in the development of a single, complete, self-consistent quantification procedure. This is uniquely different from the approach in the literature that 'quantifies' SIMS data by conventional methods (building distortions into the depth profile) and then uses deconvolution as a sharpening process.

The use of the MaxEnt method requires a model for the effects of the analysis. In the absence of an accurate process model an empirical model using the convolution integral has been provided and developed. The limitations of the model have been discussed. We have demonstrated that the instrumental response may be measured directly from delta-doped samples. We stress that this measurement should be performed concurrently with the unknown sample, as is currently considered good practice with conventional methods. The noise on a SIMS depth profile has been characterised, using MBE-grown staircase samples.

A number of algorithms for the optimisation have been assessed; the most effective to date is the simplest: the univariate search. Implicit in this algorithm is a safety feature that can recognise the symptoms of incompatible depth profile/response/noise estimate (the situation may occur due to user error). Test data has been used to determine the types of features for which conventional SIMS quantification methods are inadequate, and the performance of the MaxEnt quantification method in these cases. The

method has also been compared with inverse Fourier deconvolution (using various forms of smoothing). These comparisons were performed by carefully selecting features for which calculation of the 'depth resolution' parameters fully describes the feature.

A number of variations to the MaxEnt method have been examined, to try to improve the performance in the low-signal (background) region. The use of a novel form of the entropy which logarithmically weights the depth profile, has been found to be particularly effective.

Finally the method has been used to quantify SIMS depth profiles, separating the effects of segregation during growth from distortion during SIMS analysis, and allowing a very critical investigation of a proposed triple-delta structure.

9.2. Further Work.

The MaxEnt quantification method using a convolution integral requires a delta-doped sample in the system of interest. These are currently only routinely available for boron in silicon. As this is the 'best behaved' semiconductor system for SIMS analysis, conventional quantification methods are relatively accurate. Future development of MBE will allow the production of delta layers in other systems. This will, in turn, allow the use of MaxEnt quantification for samples where conventional methods are very poor.

The speed of the reconstruction process remains a problem: limiting the data sets to relatively few data points. Experimentally, SIMS is moving toward lower probe energy and shorter frame times. This will give far better resolution of the response function and thus an improvement in accuracy

(and possibly in noise suppression). However, it will also slow the reconstruction process to unacceptable levels. The easiest solution is to acquire faster hardware, as the rate of hardware development remains an order of magnitude at the same price every three years. This should be coupled with the adoption of a more efficient optimisation algorithm. For rapid development of the technique it was important to have code that was easily understood and adjusted. This is now much less of a limitation. Many of the more efficient optimisation codes make use of the properties of quadratic functions, which will be a particularly poor approximation to the entropy except when very close to the optimum. One of these, the *conjugate gradient* method (Fletcher and Reeves, 1964), has been implemented, but thus far has been unsuccessful. The two main problems are thought to be

(i) The scaling of the problem confounds the algorithms attempts to select a 'sensible' step size at each iteration.

(ii) The method uses interpolation between two points. This is often very unsuccessful, presumably due to the positivity constraint, as with the DSC interpolation.

Both these problems have relatively obvious solutions, but as yet the time has not been available to implement and test them.

It appears that a more accurate process model will not be available for some time: the development of the IMPETUS sputtering simulation code has not been as rapid as that of the MaxEnt method, and attempts to calculate ionisation probabilities (real or effective) from first principles have not yet been successful. Although MaxEnt quantification using the convolution model has been demonstrated to be both possible and accurate, it remains the first stage in the development of SIMS quantification methods. The ultimate goal is still to replace the model.

References.

Allen P.N, Dowsett M.G, and Collins R, *Surf. Int. Anal,* **20**, 696, (1993).

Anderson H.H, *J. Appl. Phys*, **18**, 131, (1979).

Anderson H.H, *Nucl. Instr. Meth*, **B18**, 321, (1987).

Armour D.G, Wadsworth M, Badheka R, van den Berg J.A, Blackmore G, Courtney S, Whitehouse C.R, Clark E.A, Sykes D.E, and Collins R, *Secondary Ion Mass Spectrometry, Proceedings of SIMS VI*, (eds. Benninghoven A, Huber A.M, and Werner H.W.) Wiley, Chichester, (1987).

Badheka, R, Wadsworth, M, Armour, D.G, van den Berg, J, and Clegg, J.B, *Surf. It. Anal.* **15**, 550, (1990).

Barlow R.D, Dowsett M.G, Fox H.S, Kubiak R.A.A, and Newstead S.M, *Nucl. Instr. Meth,* **B72**, 442, (1992).

Bayes T, *Philosophical Transactions of the Royal Society of London*, 330, (1763).

Benninghoven A, Rudenauer F.G, and Werner H.W. *Secondary Ion Mass Spectrometry. Basic Concepts, Instrumental Aspects, Applications and Trends*, Wiley, Chichester, (1987).

Biersack J.P, and Haggmark L.G, *Nucl. Instr. Meth.*, **174**, 257, (1980).

Brandt O, Crook G, Ploog K, Bierwolf R, Hohenstein M, Maier M, and Wagner J, *Jpn. J. Appl. Phys.* **32**, **L**24 (1993).

Brigham E.O, *The Fast Fourier Transform and its Applications*, Prentice-Hall International, London, (1988).

Carter G, Collins R, Thompson D.A, *Radiat. Effects*, **55**, 99, (1981).

Chornik B, Sopizet R, and Le Gressus C, *J. Elec. Spectr. Rel. Phenom.* **42**, 329, (1987).

Clegg J.B. and Beall, R.B, *Surf. Int. Anal.* **14**, 307, (1989).

Clegg J.B. and Beall, R.B, *Secondary Ion Mass Spectrometry, Proceedings of SIMS VII*, p.99, Wiley, Chichester (1990).

Clegg J.B. and Gale, I.G. *Surf. Int. Anal.* **17**, 190, (1991).

Collins R. and Wragg A, *J. Phys. A.* **10(9)**, 1441, (1977).

Collins R. *Radiat.Eff*, **37**, 13, (1978).

Cooke G.A, "*The development of SIMS for Two Dimensional Impurity Profiling in Semiconductors*", Ph.D. Thesis, University Of Warwick, (1992).

Cooper M. J, *Phys. Bulletin*, 463, (1977).

Cox R.P, *Am. J. Phys,* **17**, 1, (1946).

Daniell G.J, in *Maximum Entropy in Action*, p.1, Clarendon Press, Oxford, (1991).

Davidon W.C, *Math. Prog*, **9**, 1, (1975).

Dowsett M.G, and Parker E.H.C, *Int. J. Mass. Spec. and Ion Physics*, 11, (1983).

Dowsett M.G, Heal J.W, Fox H.S, and Parker E.H.C, in *Secondary Ion Mass Spectrometry, Proceedings of SIMS V*, (eds. Benninghoven A, Colton R.J, Simons D.S, and Werner H.W.), p. 176, Springer-Verlag, Berlin, (1985).

Dowsett M. G, Jeynes C, Clark E. A, Webb R, and Newstead S. M, in *Secondary Ion Mass Spectrometry, Proceedings of SIMS VII*, (eds.

Benninghoven A, Evans C.A, McKeegan K.D, Storms H.A, and Werner H.W.), p. 615, Wiley, Chichester (1990).

Dowsett M. G, Barlow R. D, Fox H. F, Kubiak R, and Collins R, *J. Vac. Sci. Technol. B* **10(1)**, 336, (1992a).

Dowsett M.G, Barlow R.D, and Allen P.N, *Proc. 2nd Int. Wkshop. on Measurement and Characterisation of Ultra Shallow Doping Profiles in Semiconductors*, **1**, 127, (1993), also *J. Vac. Sci. Technol. B*, **12(1)**, 1, (1994).

Dowsett M.G, Rowlands G, Allen P.N, and Barlow R.D, *J. Appl. Phys*, (in press).

Dowsett M.G, and Barlow R.D, *Anal. Chem. Acta.* (in press).

Eadie, W.T, Drijard, D, James, F.E, Roos, M, and Sadoulet, B, *Statistical Methods in Experimental Physics*, North-Holland Publishing, London, (1971).

Fletcher R, and Reeves C.M, *Computer J*, **7**, 149, (1964).

Gull S.F, and Daniell, G.J, *Nature*, **272**, 686, (1978).

Gull S.F, in *Maximum Entropy and Bayesian Methods*, (ed. Skilling, J.) p. 53, Kluwer, Dordrecht (1988).

Harris F.J, *Proc. IEEE*, **66(1)**, 51, (1978).

Heading J, in *Mathematical Methods in Science and Engineering*, p.611, The Chaucer Press, Suffolk, (1970).

Ho P.S. and Lewis J.E, *Surface Science*, **55**, 335, (1976).

Hofmann S, *Appl. Phys.* **9**, 59, (1976).

Hofmann S, and Sanz J.M, *Proc 8th Int. Vacuum Congress,* (Cannes,1980), vol. **1**, p.90 Thin Films.

Homma Y, and Witmark K, *J. Appl. Phys. A*, **50**, 417, (1990).

Hooke R, and Jeeves T.A, *J. Assoc. Comp. Mach*, **8**, 212, (1961).

Jaynes E.T, *Phys. Rev.* **106(4)**, 620, (1957).

Jaynes E.T, *IEEE Trans. SSC*, **4(3)**, 277, (1968).

Jesson D.E, Pennycook S.J, Baribeau J.-M, and Houghton D.C, *Thin Solid Films*, **222**, 98, (1992).

Johnson R.W, and Shore, J.E, *IEEE Trans.* **IT-29**, 942, (1983).

Jones A.F, and Misell D.L, *J. Phys. A.* **3**, 462, (1970).

King B.V, and Tsong I.S.T, *J. Vac. Sci. Technol.* A **2(4)**, 1443, (1984).

King B.V, and Tsong, I.S.T, *Nucl. Instr. Meth.* **B7/8**, 793, (1985).

Littmark U, and Hofer W.O, *Nucl. Instr. Meth.* **168**, 329, (1980).

Makarov V.V, *Surf. Int. Anal*, **20**, 821, (1993).

Mattey N.L, Dowsett M.G, Parker E.H.C, Whall T.E, Taylor S, and Zhang J.F, *Appl. Phys. Lett.* **57(16)**, 1648, (1990a).

Mattey N.L, Hopkinson M, Houghton R.F, Dowsett M.G, McPhail D.S, Whall T.E, and Parker E.H.C, *Thin Solid Films*, **184**, 15, (1990b).

McPhail D.S, Dowsett M.G, and Parker E.H.C, *Vacuum*, **36(11-12)**, 997, (1986).

Miethe K, Betz W, Nickel H, Lösch R, Schlapp W, Krimmer C, Krahn O, and Hartnagel H L, *Secondary Ion Mass Spectrometry, Proceedings of*

SIMS VIII (eds. Benninghoven A, Janssen K T F, Tümpner J, and Werner H W), p.455, Wiley, Chichester (1991).

Morgan A.E, de Grefte H.A.M, Warmoltz N, Werner H.W, and Tolle H.J, *Applications of Surface Science*, **7**, 372, (1981).

Nuttall A.H, *IEEE Trans. ASSP*, **29**, 84, (1981).

Pennycook S.J, and Jesson D.E, *Phys. Rev. Lett*, **64**, 938, (1990).

Powell A.R, Kubiak R.A.A, Whall T.E, Bowen D.K, *J. Phys. D*, **23**, 1745 (1990).

Powell A.R, Kubiak R.A.A, Whall T.E, Parker E.H.C, and Bowen D.K, *Mat. Res. Soc. Symp. Proc.* **220**, 115 (1991a).

Powell A.R, Mattey N.L, Kubiak R.A.A, Parker E.H.C, Whall T.E, and Bowen D.K, *Semicond. Sci. Technol.* **6**, 227 (1991b).

Robinson M.T, and Torrens I.M, *Phys. Rev. B*, **9(12)**, 5008, (1974).

Sanz J.M, *Surf. Int. Anal*, **6**, 196, (1984).

Savage L.J, *The Foundations of Statistics*, Wiley, New York, (1954).

Shannon C.E, and Weaver W, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, (1949).

Shore J.E, and Johnson, R.W. *IEEE Trans.* **IT-26**, 26, (1980).

Sigmund P, *Phys. Rev.* **184(2)**, 184, (1969).

Sigmund P, *J. Mat. Sci*, **8**, 1545, (1973).

Sigmund P, *J. Vac. Sci. Technol*, **A7(3)**, 585, (1989).

Sivia D.S. *Los Alamos Science*, **181**, (1990).

Skilling J, and Bryan R.K, *Mon. Not. R. Astr. Soc.* **211**, 111, (1984).

Skilling J, in *Maximum Entropy and Bayesian Methods* (ed. Skilling, J.) p. 45, Kluwer, Dordrecht (1988a).

Skilling J, in *Maximum Entropy and Bayesian methods in Science and Engineering*, Vol. 1, (eds. G.J.Erickson and C.R.Smith), p. 173, Kluwer, Dordrecht, (1988b).

Skilling J, in *Maximum Entropy in Action*, p.19 Oxford University Press, Oxford, (1991).

Spiller G D T, and Davis J R, *Secondary Ion Mass Sepectrometry, Procedings of SIMS V* (eds. Benninghoven A, Colton R J, Simons D S, and Werner H W), p.334, Springer-Verlag, Berlin (1985).

Swann W.H, in *Non-Linear Optimisation techniques*, (by Box M.J, Davies D, and Swann W.H), ICI Monograph no.5. Oliver and Boyd, Edinburgh, (1969).

Tagle J.A, Martínez Sáez V, Rojo J.M, and Salmerón M, *Surface Science* **77(1)**, 77, (1978).

Tichonov A.N, and Arsenin V.Y, *Solution of Ill-Posed Problems*. Halsted, New York, (1977).

Turner J.E, Keller H. and Mars D.E, *Secondary Ion Mass Spectrometry, Proceedings of SIMS VII*, (eds. Benninghoven, A, Evans, C.A, et al.) p.95, Wiley, Chichester (1989).

van Cittert P.H, *Z. Physik*, **69**, 298, (1931).

Voigtmann R, and Moldenhauer W, *Surf. Int. Anal*, **13**, 167, (1988).

Wernecke S.J, and d'Addario L.R, *IEEE Trans. C*, **26(4)**, 351, (1977).

Werner H.W, *Surf. Int. Anal*, **4**, 1, (1982).

Wittmark K, and Wach W, *Nucl. Instr. Meth.* **191**, 327, (1981).

Wittmark K, Dowsett M.G, and Clegg J.B, *Int. J. Mass. Spec. and Ion Physics*, **43**, 31, (1982).

Wittmark K, *Vacuum*, **34(1-2)**, 119, (1984).

Wittmark K, and Menzel N, *Appl. Phys. Lett.* **50(13)**, 815, (1987).

Yu K.S, Prutton M, Larson L.A, Pate B.B, and Poppa H,J, *Elec. Spectr. Rel. Phen.* **27**, 179, (1982).

Zalm P.C, van de Walle G.F.A, Gravesteijn, and van Gorkum A.A, *Appl. Phys. Lett*, **55(24)**, 2520, (1989).

Zalm P.C, and de Kruif R.C.M, *Appi. Surf. Sci*, **70/71**, 73, (1993).

This work was compiled using guidelines PHYS/PG/3.